

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 18:40:02 ; Search time 29.83 Seconds
(without alignments)
670.241 Million cell updates/sec

Title: US-09-828-217-1
Perfect score: 889
Sequence: 1 MASTSYDVCYRVEDGDKRC.....SSAAPQLLVLGLSLALIQ 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_032802.*
2: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
9: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
10: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
11: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
12: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
13: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
14: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
15: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
16: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
17: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
18: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
19: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
20: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
21: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	889	100.0	180	AA1980.DAT
2	889	100.0	180	AA1981.DAT
3	889	100.0	180	AA1982.DAT
4	889	100.0	180	AA1983.DAT
5	889	100.0	180	AA1984.DAT
6	889	100.0	180	AA1985.DAT
7	889	100.0	180	AA1986.DAT
8	889	100.0	180	AA1987.DAT
9	889	100.0	180	AA1988.DAT
10	889	100.0	180	AA1989.DAT
11	889	100.0	180	AA1990.DAT

12	889	100.0	193	AA1991.DAT
13	882	99.2	180	AA1992.DAT
14	843.5	94.9	197	AA1993.DAT
15	649.5	73.1	143	AA1994.DAT
16	649	73.0	132	AA1995.DAT
17	649	73.0	161	AA1996.DAT
18	579	65.1	147	AA1997.DAT
19	574.5	64.6	126	AA1998.DAT
20	473.5	53.3	155	AA1999.DAT
21	104.5	11.8	756	AA2000.DAT
22	102.5	11.5	756	AA2001.DAT
23	102.5	11.5	1484	AA2002.DAT
24	99.5	11.2	270	AA2003.DAT
25	99.5	11.2	270	AA2004.DAT
26	99.5	11.2	270	AA2005.DAT
27	99.5	11.2	273	AA2006.DAT
28	99	11.1	426	AA2007.DAT
29	99	11.1	694	AA2008.DAT
30	99	11.1	694	AA2009.DAT
31	95.5	10.7	909	AA2010.DAT
32	95.5	10.7	977	AA2011.DAT
33	95.5	10.7	977	AA2012.DAT
34	95.5	10.7	1003	AA2013.DAT
35	95.5	10.7	1240	AA2014.DAT
36	95	10.7	885	AA2015.DAT
37	94.5	10.6	111	AA2016.DAT
38	93.5	10.5	140	AA2017.DAT
39	93.5	10.5	1390	AA2018.DAT
40	93.5	10.5	1411	AA2019.DAT
41	93.5	10.5	1886	AA2020.DAT
42	93	10.5	944	AA2021.DAT
43	93	10.5	1972	AA2022.DAT
44	92.5	10.4	210	AA2023.DAT
45	92.5	10.4	22	AA2024.DAT

ALIGNMENTS

RESULT 1	
AA1980.DAT	AA1980.DAT
AA1981.DAT	AA1981.DAT
AA1982.DAT	AA1982.DAT
AA1983.DAT	AA1983.DAT
AA1984.DAT	AA1984.DAT
AA1985.DAT	AA1985.DAT
AA1986.DAT	AA1986.DAT
AA1987.DAT	AA1987.DAT
AA1988.DAT	AA1988.DAT
AA1989.DAT	AA1989.DAT
AA1990.DAT	AA1990.DAT
AA1991.DAT	AA1991.DAT
AA1992.DAT	AA1992.DAT
AA1993.DAT	AA1993.DAT
AA1994.DAT	AA1994.DAT
AA1995.DAT	AA1995.DAT
AA1996.DAT	AA1996.DAT
AA1997.DAT	AA1997.DAT
AA1998.DAT	AA1998.DAT
AA1999.DAT	AA1999.DAT
AA2000.DAT	AA2000.DAT
AA2001.DAT	AA2001.DAT
AA2002.DAT	AA2002.DAT
AA2003.DAT	AA2003.DAT
AA2004.DAT	AA2004.DAT
AA2005.DAT	AA2005.DAT
AA2006.DAT	AA2006.DAT
AA2007.DAT	AA2007.DAT
AA2008.DAT	AA2008.DAT
AA2009.DAT	AA2009.DAT
AA2010.DAT	AA2010.DAT
AA2011.DAT	AA2011.DAT
AA2012.DAT	AA2012.DAT
AA2013.DAT	AA2013.DAT
AA2014.DAT	AA2014.DAT
AA2015.DAT	AA2015.DAT
AA2016.DAT	AA2016.DAT
AA2017.DAT	AA2017.DAT
AA2018.DAT	AA2018.DAT
AA2019.DAT	AA2019.DAT
AA2020.DAT	AA2020.DAT
AA2021.DAT	AA2021.DAT
AA2022.DAT	AA2022.DAT
AA2023.DAT	AA2023.DAT
AA2024.DAT	AA2024.DAT

Human colon cancer protein bound by A protein encoded by HML1.24 antigenic p Soluble HML1.24 ant HML1.24 antigenic p HML1.24 antigenic p HML1.24 antigenic p Human HML1.24 ant Human ORFX ORF2369 Polypeptide #2 for Canine ribosome re Low density lipoprotein LDL receptor Bovine LOX-1 polyp Low density lipoprotein Human ORFX ORF1960 Human protein sequ Human transport-as Human colon cancer Amino acid sequenc Human protein SEQ Human protein SEQ Amino acid sequenc AML1 chromosome in Gene #29 human sec H. pylori GHP0 363 Drosophila melanog Nucleolar/endosoma Rattus norvegicus Novel human diagno Smooth muscle myos Bacterial general Amino acid sequenc

CC (especially 5-100) mg/kg body weight. The humanised antibody has low
 CC antigenicity and is therefore effective therapeutically in humans.
 XX
 SO Sequence 180 AA;

Query Match 100.0%; Score 889; DB 19; Length 180;
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGILVLLIIVLGVPPLIIFTIKNSACRDLRAV 60
 |||||
 Db 1 mastsydycrvpmedgdkrcklllgigilvllilvlgvppliffikansacrdglrav 60
 QY 61 MECRNVTLLQOELTEAOKGFODVFAOAATCNHTYMAIASIDAKAGOKKVEELGEI 120
 |||||
 Db 61 mecrnvthlllqgelteaqkqfvdvgaatcnhtymasidaekagqkveelgeel 120
 QY 121 TTLNHLQDASAEYERLRENOVLSVRIDAKKYYPSODSSAAAPOLLIVLGLSALLQ 180
 |||||
 Db 121 tllnhklqdaaeerlirrenqvlsvriadkkyppsgdssaaapqllivlglisallq 180

RESULT 4

AAV3202
 AAV3202 standard; Protein: 180 AA.

AAV3202;

22-NOV-1999 (first entry)

Human Hm1.24 antigenic protein.

Antigenic protein; Hm1.24; splice variant; promoter; antirheumatic;
 antiarthritic; bone marrow; tumour cell; drug development; treatment;
 myeloma; rheumatoid arthritis; human.

Homo sapiens.

WO9943803-A1.

02-SEP-1999.

25-FEB-1999; 99WO-JP00884.

25-FEB-1998; 98JP-0060617.

24-MAR-1998; 98JP-0093863.

(CHUS) CHUGAI SEIYAKU KK.

Ohtomo T, Tsuchiya M, Koishihara Y, Kosaka M;

WPI; 1999-550869/46.

N-PSDB; AAZ09726.

Genomic DNA encoding Hm1.24 antigen protein as well as splicing
 variants, useful e.g. in development of drugs for treating myeloma and
 rheumatoid arthritis

Example 1; Fig 1-2; 83pp; Japanese.

This invention describes a novel human antigenic protein, Hm1.24,

its encoding nucleic acid, splice variants and promoter region. The

products of the invention have antirheumatic and antiarthritic activity.

The DNA of the invention is isolated from bone marrow tumour cells,

which can be used to study the expression of Hm1.24 antigen, promoter

activity of its promoter region, and in development of drugs in treating

e.g. myeloma and rheumatoid arthritis. This sequence represents the

human Hm1.24 antigenic protein described in the invention.

Query Match 100.0%; Score 889; DB 20; Length 180;
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGILVLLIIVLGVPPLIIFTIKNSACRDLRAV 60
 |||||
 Db 1 mastsydycrvpmedgdkrcklllgigilvllilvlgvppliffikansacrdglrav 60
 QY 61 MECRNVTLLQOELTEAOKGFODVFAOAATCNHTYMAIASIDAKAGOKKVEELGEI 120
 |||||
 Db 61 mecrnvthlllqgelteaqkqfvdvgaatcnhtymasidaekagqkveelgeel 120
 QY 121 TTLNHLQDASAEYERLRENOVLSVRIDAKKYYPSODSSAAAPOLLIVLGLSALLQ 180
 |||||
 Db 121 tllnhklqdaaeerlirrenqvlsvriadkkyppsgdssaaapqllivlglisallq 180

RESULT 5

AAV32765
 AAV32765 standard; Protein: 180 AA.

AAV32765;

25-OCT-1999 (first entry)

Soluble Hm1.24 antigenic protein sequence.

Hm1.24; antigenic protein; antibody; immunoassay; diagnosis; cancer;

immune disorder; immunotherapy monitoring.

Homo sapiens.

WO9943703-A1.

02-SEP-1999.

25-FEB-1999; 99WO-JP00885.

25-FEB-1998; 98JP-0060613.

(CHUS) CHUGAI SEIYAKU KK.

Koishihara Y, Ozaki Y;

WPI; 1999-518836/43.

N-PSDB; AAZ10917.

Immunoassay of anti-Hm1.24 antibody or soluble Hm1.24 antigen,
 useful for diagnosis of immune disorders and cancer

Example 3; Fig 14-15; 138pp; Japanese.

This sequence represents a human soluble Hm1.24 antigenic protein. The
 invention relates to an immunochemical assay of anti-Hm1.24 antibody by
 use of a soluble Hm1.24 antigenic protein, or an immunochemical assay of
 the soluble antigen by use of the antibody. The immunoassay of the Hm1.24
 antigen or antibody is useful for diagnosis of immune disorders and
 cancer, for monitoring of anti-Hm1.24 antibody immunotherapy, and for
 assay of the antibody or antigen for investigative purposes, in
 CC biological samples such as blood, serum, urine, milk, synovial fluid or
 CC microorganism culture media. The method is sensitive down to 500 pg/ml
 CC antibody.

Sequence 180 AA;

Query Match 100.0%; Score 889; DB 20; Length 180;
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGILVLLIIVLGVPPLIIFTIKNSACRDLRAV 60
 |||||
 Db 1 mastsydycrvpmedgdkrcklllgigilvllilvlgvppliffikansacrdglrav 60

OY		61	MECRWVTHLLOQLTEAOKGFODVEAOATCNHTWALMASIDAERKAGQKVEELEGEI	120
Dd		61	mecrnvthllqqltleaqqkfqvdeaaatcnhtwalmasidaekagqkvveelgeei	120
OY		121	TTLNHLKLDASAEVERLRRENOVL SVRIADKKRYPSQDSSAAAPOLLIVLLIGLSALLQ	180
Dd		121	tllnhklldasaeveerlrrengvlsvriadkkrypsqdssaaaapqlllvllglsallq	180
 RESULT 6 RAY02576 ID AAY02576 standard; Protein: 180 AA. XX AAY02576; 16-JUL-1999 (first entry) HML_24 antigenic protein. Reconstituted human antibody; peptide antigen HML_24; framework region; KW complementary determining region; CDR; anti-HML_24 antibody; myeloma; KH humanised antibody. XX OS Homo sapiens. XX PN MO9918212-A1. XX PD 15-APR-1999. XX XX 02-OCT-1998; 98WO-JP04469. XX PR 03-OCT-1997; 97JP-0271726. XX PA (CHUS) CHUGAI SEITYAKU KK. XX PI Tsuchiya M; XX DR WPL: 1999-277273/23. XX DR N-PSDB: AAX59485. PT Reconstituted human antibody useful in the treatment of myeloma				
 Disclosure; Page 131-133; 256pp; Japanese.				
CC	The specification describes a reconstituted human antibody recognizing			
CC	the peptide antigen HML_24. This human antibody contains natural human			
CC	framework regions modified by amino acid substitutions to provide			
CC	a homogeneity with a previously designed framework region (which may			
CC	arise from a human or non-human source); and complementary determining			
CC	regions (CDR) derived from a non-human anti-HML_24 antibody. The			
CC	reconstituted antibody is useful in the treatment of diseases in which			
CC	the surface antigen HML_24 is implicated such as myeloma. The present			
CC	sequence represents HML_24 antigenic protein.			
SQ	Sequence 180 AA;			
 Query Match 100.0%; Score 889; DB 20; Length 180; Best Local Similarity 100.0%; Pred. No. 6.2e-78; Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	1	MASTSYDYCRVPMEGDKRKCLLIGILVLLITVILGVPLITFTIKANSEACRGDLRAV	60	
Dd	1	mastysdyrcrvpmedgdkrkcllligilvlllitvllgvplllftikanseacrgdlrav	60	
OY	61	MECRWVTHLLOQLTEAOKGFODVEAOATCNHTWALMASIDAERKAGQKVEELEGEI	120	
Dd	61	mecrnvthllqqltleaqqkfqvdeaaatcnhtwalmasidaekagqkvveelgeei	120	
OY	121	TTLNHLKLDASAEVERLRRENOVL SVRIADKKRYPSQDSSAAAPOLLIVLLIGLSALLQ	180	
Dd	121	tllnhklldasaeveerlrrengvlsvriadkkrypsqdssaaaapqlllvllglsallq	180	

RESULT	7
AA05484	
ID	AA05484 standard; Protein; 180 AA.
XX	
AC	AA05484;
XX	
DT	07-JUL-1999 (first entry)
XX	
DE	Potentiator for antibody against lymphoid tumour.
XX	
KW	Antibody potentiator; lymphoid tumour; lymphoma; cytotoxic antibody;
KW	multiple myeloma; acute B-lymphocytoma; chronic B-lymphocytoma;
KW	pre-B lymphoma; Burkitt's lymphoma; acute T-lymphocytoma; therapy;
KW	chronic T-lymphocytoma; PNTL.
XX	
OS	Homo sapiens.
XX	
PN	WO9918997-A1.
XX	
PD	22-APR-1999.
XX	
PE	14-OCT-1998; 98WO-JP04645.
XX	
PR	05-AUG-1998; 98JP-0222024.
PR	14-OCT-1997; 97JP-0280759.
XX	
PA	(CHUS) CHUGAI SEIYAKU KK.
XX	
PI	Koishihara Y, Kosaka M;
DR	WPI: 1999-277447/23.
DR	N-PSDB: AAX36561.
XX	
PT	Potentiation of antibody treatment of lymphoma with biological
PT	response modifier
XX	
PS	Claim 1; Page 37-38; 62pp; Japanese.
XX	
CC	This sequence represents a potentiator for an antibody against
CC	lymphoid tumour.
CC	The invention relates to a method for the treatment of lymphoma, in which
CC	a cytotoxic antibody is potentiated by administration of a biological and
CC	response modifier. The method can be used for treatment of lymphomas and
CC	multiple myelomas which are resistant to conventional treatment, such as
CC	acute B-lymphocytoma, chronic B-lymphocytoma, pre-B lymphoma, Burkitt's
CC	lymphoma, acute T-lymphocytoma, chronic T-lymphocytoma, and PNTL.
XX	
SO	Sequence 180 AA;
Query Match	100.0%; Score 889; DB 20; Length 180;
Best Local Similarity	100.0%; Pred. No. 6,2e+78;
Matches 180; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	1 MASTSYDVCRAVMEGGDRCKLLIGILIVLLITVIGLVPLIIFTTKANSEACROSLRAY 60
Db	1 mastsydyrcrvmedgdkricklllgilivllitvlligvpilliftkansseacrdglrav 60
OY	61 MECAVNTHILOELTEAOKGFODVEAOATCNHTVMAALMASLDAEKAQOKKVEEEGET 120
Db	61 mecrnvthlqlgelteaqdgifgdveagaatcnhtymalnasidaekagqkkveeleget 120
OY	121 TTLNHRLODASAEVRRLRRNOVLVSRIADKKYYRSSQDSSNAAPQLILVILGLSALIQ 180
Db	121 tlinhkldaseverlrrenvgvlsvriadkkyrpsqssaaapqllilvilglsaliq 180
RESULT	8
AA07250	
ID	AA07250 standard; Protein; 180 AA.
XX	

AC AAY07250;
 XX
 DE 06-JUL-1999 (first entry)
 XX
 DE BST-2 protein.
 XX
 KM Mouse; BST-2; monoclonal antibody; RS38; myeloma; cytotoxic activity.
 XX
 OS Mus sp.
 XX
 PN JP11092399-A.
 XX
 PD 06-APR-1999.
 XX
 PF 24-SEP-1997; 97JP-0274960.
 XX
 PR 24-SEP-1997; 97JP-0274960.
 XX
 PA (CHUS) CHUGAI PHARM CO LTD.
 XX
 DR WPI: 1999-283503/24.
 XX
 DR N-PSDB; AAX29996.
 XX
 PT An agent for treating myeloma -includes an antibody and has
 XX
 AS cytotoxic activity
 XX
 CS Claim 1; Page 10; 13pp; Japanese.
 XX
 CC This sequence represents the mouse BST-2 protein which is used to raise
 CC antibodies, especially the monoclonal antibody RS38. The antibody can be
 CC used in compositions to treat myelomas when the antibody is associated
 CC with a cytotoxic activity.
 XX
 SQ Sequence 180 AA;

Query Match 100.0%; Score 889; DB 20; Length 180;
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDCRYPMEDGDKRCKLLIGILVLLIIVIGVPLIIFTIKANSFACDGLRAV 60
 DB 1 mastsydcrypmmedgdkrcrklllgilvllivllivpliftikansacrdglrav 60
 QY 61 MECHNVTHLLQOELTEAOKGFQDVEAQAATCNHTVMALMASLDAEKAQGGKVEELGEI 120
 DB 61 mecnvthlllqgelteaqkgfqedveaqaatcnhtvmalmasldeakagqgkveelgei 120
 QY 121 TTLNHLKQDASAEVERLRRENOVLSVRIADKKYPPSSODSSSAAPOLLIVLGLSALLQ 180
 DB 121 ttlhnlkqdasaeverlrrenovlsvriadkkypssodsssaapqllivlglisallq 180

RESULT 9
 ID AAY53273
 AC AAY53273 standard; Protein; 180 AA.

AC AAY53273;
 XX
 DT 21-JUL-2000 (first entry)
 XX
 DE Human Hm1.24 antigen protein sequence SEQ ID NO:2.
 XX
 KM Human; Hm1.24 antigen protein; detection; plasmocytoma;
 KM multiple myeloma; plasmocytic leukaemia; extramedullary plasmocytoma;
 KM multiple plasmocytoma; asymptomatic myeloma.
 XX
 OS Homo sapiens.
 XX
 PN WO200017395-A1.
 XX
 PD 30-MAR-2000.
 XX

PF 20-AUG-1999; 99WO-JP04502.
 XX
 PR 18-SEP-1998; 98JP-0264593.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 XX
 PI Kawai S, Koishibara Y, Kosaka M;
 XX
 DR WPI: 2000-283616/24.
 XX
 DR N-PSDB; AAA13654.
 XX
 PT Detection or measurement of plasmocytomas, applicable for early
 PT diagnosis of e.g. multiple myeloma and plasmocytic leukemia, using a
 PT polynucleotide which is expressed specifically or strongly in
 XX
 PS plasmocytomas
 XX
 PS Disclosure; Page 15-16; 20pp; Japanese.
 XX
 CC A method has been developed for detecting or measuring plasmocytomas in
 CC a sample at an early stage of disease development. The method comprises
 CC amplifying a polynucleotide which is expressed specifically or strongly
 CC in plasmocytomas before quantifying the amplification product by
 CC comparing with results obtained with a control sample. The method is for
 CC detecting or measuring plasmocytomas, applicable for early diagnosis of
 CC e.g. multiple myeloma, plasmocytic leukaemia, isolated plasmocytoma,
 CC extramedullary plasmocytoma, multiple plasmocytoma obtained from smoking
 CC or asymptomatic myeloma. The present sequence represents human Hm1.24
 CC antigen protein, which is expressed in plasmocytomas and so can be used
 CC in the method of the invention.
 XX
 SQ Sequence 180 AA;

Query Match 100.0%; Score 889; DB 21; Length 180;
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDCRYPMEDGDKRCKLLIGILVLLIIVIGVPLIIFTIKANSFACDGLRAV 60
 DB 1 mastsydcrypmmedgdkrcrklllgilvllivllivpliftikansacrdglrav 60
 QY 61 MECHNVTHLLQOELTEAOKGFQDVEAQAATCNHTVMALMASLDAEKAQGGKVEELGEI 120
 DB 61 mecnvthlllqgelteaqkgfqedveaqaatcnhtvmalmasldeakagqgkveelgei 120
 QY 121 TTLNHLKQDASAEVERLRRENOVLSVRIADKKYPPSSODSSSAAPOLLIVLGLSALLQ 180
 DB 121 ttlhnlkqdasaeverlrrenovlsvriadkkypssodsssaapqllivlglisallq 180

RESULT 10
 ID ABB50295
 AC ABB50295 standard; Protein; 180 AA.

AC ABB50295;
 XX
 DT 08-FEB-2002 (first entry)
 XX
 DE Bone marrow stromal antigen (BST-2) ovarian tumour marker protein, #80.
 XX
 KM Ovarian tumour marker gene; human; overexpression; upregulation;
 KM epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
 KM identification; serous cystadenoma; borderline serous tumour;
 KM serous cystadenocarcinoma; mucinous cystadenocarcinoma;
 KM mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
 KM undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
 KM adenocarcinoma; Brenner tumour; serial analysis of gene expression; SAGE;
 KM immune response pathway; cell proliferation regulation; protein folding;
 KM membrane localised; secreted; therapeutic target; cytosstatic;
 KM gene therapy; vaccine.
 XX
 OS Homo sapiens.
 XX

AC AAG73947;
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:4711.
DE
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 19.
XX
XX Homo sapiens.
OS
XX WO200122920-A2.
PN
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000WO-US26524.
PF
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
XX N-PSDB; AAH33378.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 11; Page 6512-6513; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 193 AA:
Query Match 100.0%; Score 889; DB 22; Length 193;
Best Local Similarity 100.0%; Pred. No. 6.8e-78;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASTSYDYCRVPMEDGDRCKLLGIGILVLLIIVLGVPILITFIKANSEACRDLRAV 60
DB 14 mastsydycrvpmegdgrcklllgigilvlllilvlgvpiliflikanseacrdglrav 73
QY 61 MECRWVTHLLOELTEAOKGFQDVEAQAATCNHTVMALMASIDAKAGOKKVEELEGEI 120
DB 74 mecrwvthllqgelteaokgfgdvgaatcnhtvmalmasidaekagqkkveelegei 133
QY 121 TTLNKKLDASAEVRLRENOVLSVRADKKYRSPSSODSSAAAPOLLIVILGSLALQ 180
DB 134 tllnkkldasaevevrlrenovlsvriadkkkyrpsqdssaaapqllivlilglsallq 193
RESULT 13

AAW77292
ID AAW77292 standard; Protein; 180 AA.
XX
XX AAW77292;
AC
XX 14-DEC-1998 (first entry)
DE
XX Protein bound by Anti-HM1.24 antibody.
DE
XX Anti-HM1.24; antibody; lymphocyte activation inhibitor;
KW autoimmune disease; organ transplant; allergy.
XX
XX Homo sapiens.
OS
XX WO9837913-A1.
PN
XX 03-SEP-1998.
PD
XX 27-FEB-1998; 98WO-JP00831.
PF
XX 28-FEB-1997; 97JP-0045663.
PR
XX (CHUS) CHUGAI SEIYAKU KK.
PA
XX Koishihara Y;
PI WPI: 1998-480937/41.
XX N-PSDB; AAW59114.
XX
XX Lymphocyte activation inhibitor comprises antibodies, particularly
PT anti-HM1.24 antibody - for preventing and treating autoimmune
XX diseases, rejection reactions in organ transplant or allergy
XX
XX Disclosure: Page 38-39; 53pp; Japanese.
XX
XX The Anti-HM1.24 antibody can be used in the production of lymphocyte
CC activation inhibitors. These inhibitors can be used for the prevention
CC and treatment of autoimmune diseases, rejection reactions in organ
CC transplant or allergy. Administration is non-oral, e.g. by intra
CC venous and intramuscular injection, local or systemic.
XX
XX Sequence 180 AA:
Query Match 99.2%; Score 882; DB 19; Length 180;
Best Local Similarity 99.4%; Pred. No. 3e-77;
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MASTSYDYCRVPMEDGDRCKLLGIGILVLLIIVLGVPILITFIKANSEACRDLRAV 60
DB 1 mastsydycrvpmegdgrcklllgigilvlllilvlgvpiliflikanseacrdglrav 60
QY 61 MECRWVTHLLOELTEAOKGFQDVEAQAATCNHTVMALMASIDAKAGOKKVEELEGEI 120
DB 61 mecrwvthllqgelteaokgfgdvgaatcnhtvmalmasidaekagqkkveelegei 120
QY 121 TTLNKKLDASAEVRLRENOVLSVRADKKYRSPSSODSSAAAPOLLIVILGSLALQ 180
DB 121 tllnkkldasaevevrlrenovlsvriadkkkyrpsqdssaaapqllivlilglsallq 180
RESULT 14
AAW36951
ID AAW36951 standard; Protein; 197 AA.
XX
XX AAW36951;
AC
XX 12-MAY-1998 (first entry)
DE
XX Protein encoded by clone 0238_1.
DE
XX Human; secreted protein; molecular weight marker; genetic fingerprinting;
KW antibody production; nutritional supplement; therapy; clone 0238_1;
XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 18:42:17 ; Search time 27.05 Seconds
(without alignments)
1151.168 Million cell updates/sec

Title: US-09-828-217-1
Perfect score: 889
Sequence: 1 MASTSYDYGCRVPMEDGDKRC.....SSAAPQLLIVLLGLSALLQ 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	12.5	778	12 O9YMP5	O9YMP5 lymantria d
2	107.5	12.1	461	16 O9K6X4	O9K6X4 bacillus ha
3	107	12.0	143	16 O9KA43	O9KA43 bacillus ha
4	104.5	11.8	756	4 O9Y6W2	O9Y6W2 homo sapien
5	104.5	11.8	756	4 O9NRK8	O9NRK8 homo sapien
6	104.5	11.8	756	4 O9Y6W1	O9Y6W1 homo sapien
7	103.5	11.6	782	4 O9NKK3	O9NKK3 homo sapien
8	103.5	11.6	782	4 O9NKK4	O9NKK4 homo sapien
9	102.5	11.5	1534	6 O28298	O28298 canis famli
10	102.5	11.5	1627	5 O96200	O96200 gliardia lam
11	101.5	11.4	304	11 O91Y73	O91Y73 mus musculu
12	101.5	11.4	826	13 O9YHD5	O9YHD5 rana catesb
13	100.5	11.3	506	4 O9NWT9	O9NWT9 homo sapien
14	100	11.2	865	13 O98SN6	O98SN6 gallus gall
15	99.5	11.2	270	6 P79391	P79391 bos taurus
16	99.5	11.2	708	13 O9YHD7	O9YHD7 rana catesb

17	99.5	11.2	1932	13 O98TQ4	O98TQ4 notothenia
18	99	11.1	415	2 O54859	O54859 streptococc
19	99	11.1	800	3 O96X03	O96X03 emeticocella
20	98.5	11.1	712	13 O98SN5	O98SN5 gallus gall
21	98.5	11.1	975	13 O98TQ5	O98TQ5 notothenia
22	98.5	11.1	1931	13 O42352	O42352 cyprinus ca
23	98.5	11.1	1936	13 O90YF6	O90YF6 paracitirhit
24	98.5	11.1	1963	5 O02244	O02244 caenorhabdit
25	97.5	11.0	1938	6 O9GYP9	O9GYP9 oryctolagus
26	97	10.9	425	2 P95808	P95808 streptococc
27	97	10.9	1937	6 O9TV62	O9TV62 sus scrofa
28	97	10.9	1941	5 O26079	O26079 piacopecten
29	97	10.9	1950	5 O26080	O26080 piacopecten
30	97	10.9	2138	5 O9XZ83	O9XZ83 amoeba prot
31	96.5	10.9	1933	13 O90337	O90337 cyprinus ca
32	96	10.8	761	13 P79793	P79793 gallus gall
33	95.5	10.7	173	16 O55887	O55887 synechocyst
34	95.5	10.7	360	2 P70906	P70906 borrelia he
35	95.5	10.7	671	13 O9YHD4	O9YHD4 rana catesb
36	95.5	10.7	977	4 O75300	O75300 homo sapien
37	95.5	10.7	977	4 O9H476	O9H476 homo sapien
38	95.5	10.7	1341	12 O88304	O88304 sandfly fev
39	95.5	10.7	1407	4 O96S82	O96S82 homo sapien
40	95.5	10.7	1586	4 O9P2E9	O9P2E9 homo sapien
41	94.5	10.6	209	13 O9PYE1	O9PYE1 brachydanio
42	94.5	10.6	682	10 O94GH0	O94GH0 oryza sativ
43	94.5	10.6	870	11 O9CS20	O9CS20 mus musculu
44	94.5	10.6	1092	13 O90338	O90338 cyprinus ca
45	94.5	10.6	1388	11 P70336	P70336 mus musculu

ALIGNMENTS

RESULT 1
O9YMP5 PRELIMINARY; PRT; 778 AA.
ID O9YMP5
AC O9YMP5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IDORF-82 PEPTIDE.
OS Lymantria dispar multicaud nuclear polyhedrosis virus (LdMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
CC Nucleopolyhedrovirus.
OX NCBI_TaxID=10449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99124785; PubMed=9887315;
RA Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
RA Slavicek J.M., Rohmann G.F.;
RT "Sequence and analysis of the genome of a baculovirus pathogenic for
RT Lymantria dispar."
RL Virology 253:17-34(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Kuzio J.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF081810; AAC70268.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; protamine_P1; 1.
SQ SEQUENCE 778 AA: 87841 MW: 81188BD172CE9E43 CRC64;

Query Match 12.5%; Score 111; DB 12; Length 778;
Best Local Similarity 32.5%; Pred. No. 0.4;
Matches 38; Conservative 22; Mismatches 47; Indels 10; Gaps 4;

OY 46 IKANSEACRDGLRAYMECR-NWTHLLQDELTAQKGFQDVEQAATCNITWALASIDA 104
DB 520 IKAQSELNMD-LQAKAEQADANARLQAEISLKR---AESDADLRRVVAQLAEAES 574
OY 105 EKAQCGKKEVELEGEITTLNHLQDASAEVERLRENGVLSVRIDKKYPPSSQSS 161

Db	575	GAADLQNRIRRLAEASSGILTRRLQESAAEVLALRDKREDLERRTAAS----	AAQDVS	627
RESULT	2			
ID	Q9K6X4	PRELIMINARY:	PRT:	461 AA.
AC	Q9K6X4			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	CELL WALL-BINDING PROTEIN.			
OS	BH3600.			
OS	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=86665;			
SP	11]			
SEQUENCE	FROM N.A.			
STRAIN	C-125 / JCM 9153;			
MEDLINE	=20512582; PubMed=11058132;			
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,			
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,			
RA	Horikoshi K.;			
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus			
RL	halodurans and genomic sequence comparison with Bacillus subtilis.";			
RL	Nucleic Acids Res. 28:4317-4331(2000).			
DR	EMBL: AP001519; BAB07319.1; "			
DR	MEROPS: M37.0PW; "			
DR	Interpro: IPR002886; Peptidase_M37.			
DR	Pfam: PF01551; Peptidase_M37; 1.			
KM	Complete proteome.			
SO	SEQUENCE 461 AA; 50372 MW; 2918480CD67AF3F CRC64;			
Query Match	12.1%; Score 107.5; DB 16; Length 461;			
Best Local Similarity	26.1%; Pred. No. 0.44;			
Matches	31; Conservative 28; Mismatches 55; Indels 5; Gaps			
OY	33 IYILGVPLIITFIKANSACRDLRAVWECRNTHLIQELTEAOKFQDVEAQAATCN 922			
Db	5 ISLVAAAGLITFSLITSSSIDEA-KANSSLDNQISDVQKEROEKQOKTEAL----			59
OY	93 HTVVALASADAKAGOGKVEELGETTLNHHKLDASAVEFLRENOVLSVRIADK 151			
Db	60 KEVEKELGDTTAEITAEIKLKEVEETSGKIOEKREELIEVQAEITELKQELIERIAER 118			
SOUT	3			
ID	Q9KAA3	PRELIMINARY:	PRT:	143 AA.
AC	Q9KAA3			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	FLAGELLAR PROTEIN REQUIRED FOR FLAGELLAR FORMATION.			
OS	FTIL OR BH2447.			
OS	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=86665;			
SP	11]			
SEQUENCE	FROM N.A.			
STRAIN	C-125 / JCM 9153;			
MEDLINE	=20512582; PubMed=11058132;			
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,			
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,			
RA	Horikoshi K.;			
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus			
RL	halodurans and genomic sequence comparison with Bacillus subtilis.";			
RL	Nucleic Acids Res. 28:4317-4331(2000).			
DR	EMBL: AP001515; BAB06166.1; "			
KM	Flagella; Complete proteome.			

Query Match	12.0%	Score 107	DB 16	Length 143
Best Local Similarity	22.1%	Pred. No. 0.13		
Matches	31	Conservative	34	Mismatches 49; Indels 26; Gaps 4
QY	23	LGIGLIVLLIVILVGLVLIIFTIKANSEACRGLRAVMEC-----RNVTHLLOEL	74	
Db	6	LVNIMLILIVLTIVTGVAVLIVFNPNNEDEQDREPTIDEIIAQSVEPEEITNLSNDF	65	
QY	75	TEA-----QKGQDYDAQATCNHTVYVLMASLIDEKAQOGKVELESEITTLNKK	126	
Db	66	VRARFLIHDVNNMALQEVOKRDFQVYNNIIIRSLAGMDASQLSGAGIEXLEAO-----	118	
QY	127	LQDASAEVERLRRENOVLVS	146	
Db	119	LQD-----DINALMGESVYKI	135	
RESULT	4			
Q9Y6W2	09Y6W2	PRELIMINARY;	PRT;	756 AA.
AC	Q9Y6W2:			
DT	01-NOV-1999	(TREMBLrel. 12. Created)		
DT	01-NOV-1999	(TREMBLrel. 12. Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19. Last annotation update)		
DE	HC	PROTEIN.		
GN	HC			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid:9606;			
RA	SEQUENCE FROM N.A.			
RX	MEDLINE=20014706; PubMed=10545595;			
RA	Oka A., Tamiya G., Tomizawa M., Ota M., Katsuyama Y., Makino S.,			
RA	Shima T., Yoshitome M., Lizuka M., Saso Y., Iwashita K.,			
RA	Kavakudo Y., Sugai J., Ozawa A., Ohkido M., Kimura M., Bahram S.,			
RA	Inoko H.;			
RT	"Association analysis using refined microsatellite markers localizes a			
RT	susceptibility locus for psoriasis vulgaris within a 11kb segment			
RT	telomeric to the HLA-C gene."			
RL	Hum. Mol. Genet. 8:2165-2170(1999).			
DR	EMBL, AB029331; BAA81890.1;..			
SO	SEQUENCE	756 AA;	86118 MW;	88PDSF858EF07601 CRC64;
Query Match	11.8%	Score 104.5;	DB 4;	Length 756;
Best Local Similarity	24.7%	Pred. No. 1.4;		
Matches	39;	Conservative	34;	Mismatches 48; Indels 37; Gaps 7;
QY	54	RDGLRAVME-----CRNVTHLI--QQLTEAQKGFQVDEAO-AAATC-----NHTYWA	97	
Db	258	RDSLHATVTELLQVRRVQSLTHIALQEEELTRKVPSPDSLEPEFTKKCOSLIRNREKVFA	317	
QY	98	LMASLDKAKAGQKQKVELEGEITLNNK-----LQDASAEVERLRRENOV	143	
Db	318	LMVQIKAELEHSDSVSKQKQGVASLQEKVTSQSQEQAILORSLODKRAEVEVERMGAKG	377	
QY	144	LSVRIA---DKKYPSSQDSSSAAPQLLIVLIGLSA	177	
Db	378	LQLELSRAQEARRMW---QQQTASAEEDQLRLVNVAVSS	412	
RESULT	5			
Q9NRK8	Q9NRK8	PRELIMINARY;	PRT;	756 AA.
AC	Q9NRK8:			
DT	01-OCT-2000	(TREMBLrel. 15. Created)		
DT	01-OCT-2000	(TREMBLrel. 15. Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19. Last annotation update)		
DE	A-HELICAL	PROTEIN.		

GN HCR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347693; PubMed=10888604;
RA Asumelahti K., Laitinen T., Ilkonen-Vatjus R., Lohi M.-L.,
RA Suomela S., Snellman E., Saarialho-Kere U., Kere J.,
RT "A candidate gene for psoriasis near HLA-C, HCR (Pg8), is highly
RT polymorphic with a disease-associated susceptibility allele."
RL Hum. Mol. Genet. 9:1533-1542(2000).
DR EMBL: AF216493; AAF74221.1;
SQ SEQUENCE 756 AA; 86058 MW; 8E4D03358B62DEB4 CRC64;

Query Match 11.8%; Score 104.5; DB 4; Length 756;
Best Local Similarity 24.7%; Pred. No. 1.4;
Matches 39; Conservative 34; Mismatches 48; Indels 37; Gaps 7;

54 RDGLRAVME-----CRNVTLL-----QDELTEAKGFQDVEAO-AATC-----NHTVMA 97
Db 258 RDSLHATFAELLQVRVQSLTHIALQEEELTRKVPDSLEPEFRKCCSLNRMREKVFA 317
OY 98 LMSLDAKAKGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143
Db 318 LMSLDAKAKGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 317
OY 144 LSVRIA-----DKKYPSSODSSSAAPOLLIVLGLSA 177
Db 378 LQELSLRAQAEARRRW-----QQQTASAEQLRLVNAVSS 412

RESULT 6
OY6W1 PRELIMINARY; PRT; 756 AA.
AC OY6W1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HCR PROTEIN (TRICOMYALIN HOMOLOGUE).
GN HCR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=BLOOD;
RA Oka A., Tamai G., Makino S., Tomizawa M., Yamagata T., Shilina T.,
RA Watanabe K., Yamazaki M., Tashiro H., Kimura M., Inoko H.,
RT "HCR-a-helical coiled-coil rod homologue."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hirakawa M., Yamaguchi H., Imai K., Shimada J.,
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shilina S., Tamai G., Oka A., Inoko H.,
RT "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB029343; BAA82158.1;
DR EMBL: AP000509; BAB63313.1;
SQ SEQUENCE 756 AA; 85951 MW; 3D109AFOEEFAB9P5 CRC64;

Query Match 11.8%; Score 104.5; DB 4; Length 756;
Best Local Similarity 25.2%; Pred. No. 1.4;
Matches 40; Conservative 32; Mismatches 48; Indels 39; Gaps 7;
OY 54 RDGLRAVME-----CRNVTLL-----QDELTEAKGFQDVEAO-AATC-----NHTVMA 97

Db 258 RDSLHATFAELLQVRVQSLTHIALQEEELTRKVPDSLEPEFRKCCSLNRMREKVFA 317
OY 98 LMSLDAKAKGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143
Db 318 LMSLDAKAKGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 317
OY 144 LSVRIA-----DKKYPSSODSSSAAPOLLIVLGLSA 177
Db 378 LQELSLRAQAEARRRW-----QQQTASAEQLRLVNAVSS 412

RESULT 7
OY6W1 PRELIMINARY; PRT; 782 AA.
AC OY6W1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE CDNA FLJ20197 FIS, CLONE COLF0996.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono S., Sugano S.,
RT "NEDO human CDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000204; BAA91007.1;
SQ SEQUENCE 782 AA; 86616 MW; F472FE544F627CE8 CRC64;

Query Match 11.6%; Score 103.5; DB 4; Length 782;
Best Local Similarity 25.2%; Pred. No. 1.7;
Matches 40; Conservative 32; Mismatches 48; Indels 39; Gaps 7;
OY 54 RDGLRAVME-----CRNVTLL-----QDELTEAKGFQDVEAO-AATC-----NHTVMA 97
Db 284 RDSLHATFAELLQVRVQSLTHIALQEEELTRKVPDSLEPEFRKCCSLNRMREKVFA 343
OY 98 LMSLDAKAKGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143
Db 344 LMSLDAKAKGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 343
OY 144 LSVRIA-----DKKYPSSODSSSAAPOLLIVLGLSA 177
Db 404 LQELSLRAQAEARRRW-----QQQTASAEQLRLVNAVSS 438

RESULT 8
OY6W1 PRELIMINARY; PRT; 782 AA.
AC OY6W1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CDNA FLJ20210 FIS, CLONE COLF1787.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono S., Sugano S.,
RT "NEDO human CDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AK000217; BAA91016.1; -
SQ SEQUENCE 782 AA; 88643 MW; A510C8BCFA8247B CRC64;

Query Match
Best Local Similarity 25.2%; Pred. No. 1.7;
Matches 40; Conservative 32; Mismatches 48; Indels 39; Gaps 7;

QY 54 RDGLRAVME-----CRNVTHLT---OOELTEAOKGFQDVEAQAATCNHTVMALMSL-DA 104
DB 284 RDSLHATTAELLQVRVSLTHLTLAOLGEELTRKYVPSDSEPEPTRKQSLNFMREKVF 343
QY 98 LMSLDAEKAQOKKVEELEGETTINHK-----LQDASAEVELRRENOY 143
DB 344 LMVQLAKAELEHSDSVKQKGVASLOEKVTSQSOQAILQRSLODKAAVEVERKANG 403
QY 144 LSVRI-----ADKKYPPSSODSSSAAPOLLIVLGLISA 177
DB 404 LOELSRAGEARROW-----QQQTASAEQLRLVYNAVSS 438

RESULT 9

ID 028298 PRELIMINARY; PRT; 1534 AA.
AC 028298;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RIBOSOME RECEPTOR.
GN P180.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95310363; PubMed=7790375;
RA Wanner E.E., Sun Y., Savitz A.J., Meyer D.I.;
RT "Functional characterization of the 180 kDa ribosome receptor in vivo."
RL J. Cell Biol. 130:29-39(1995).
DR EMBL: X87224; CAA60676.1; -
KW Receptor.
SQ SEQUENCE 1534 AA; 164586 MW; B343BCF12656F3C5 CRC64;

Query Match
Best Local Similarity 11.5%; Score 102.5; DB 6; Length 1534;

Matches 26; Conservative 22; Mismatches 42; Indels 1; Gaps 1;

QY 55 DGLRAVMECRNVTHLTLOELTEAOKGFQDVEAQAATCNHTVMALMSLDAEKAQOKKVE 114
DB 847 DAAVASKRLREVKELAAEKAAGAKYKQLVAREGETFVQRIEASVREHYKEVQ 906
QY 115 ELEGETTINHKIQDA-SAEVELRRENOY 144
DB 907 QLOGKIRTLQEOLENGPNTOLARLQOENSIL 937

RESULT 10

ID 096200 PRELIMINARY; PRT; 1627 AA.
AC 096200;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AXONEME-ASSOCIATED PROTEIN GASP-180.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RA Elmendorf H.G., Rohrer S.C., La Vigne E.A., Nash T.E.;

RT "Novel Axoneme-Associated Proteins in Giardia lamblia."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF400249; AAK91740.1; -
SQ SEQUENCE 1627 AA; 179270 MW; 021EED9763907DCC CRC64;

Query Match
Best Local Similarity 11.5%; Score 102.5; DB 5; Length 1627;
Matches 34; Conservative 23; Mismatches 60; Indels 5; Gaps 3;

QY 46 IKANSACRDGLRAVMECRNVTHLTLOELTEAOKGFQDVEAQAATCNHTVMALMSL-DA 104
DB 1284 LRESAELQDCKLHAUSDSDRADGDLOKLYQLEKDISGAKELVAERDATTDELKORLRT 1343
QY 105 EKAQOKK-RVEELEGETTINHKIQDASAEVELRRENOYLSRIADKKYPPSSODSSSA 163
DB 1344 EEVDLKEKHAELDEIDIVNLNGLKKDKDELIELRLRQ---LEAQPTATVYVESGSEVGD 1400
QY 164 AA 165
DB 1401 AA 1402

RESULT 11

ID 091Y73 PRELIMINARY; PRT; 304 AA.
AC 091Y73;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MACROPHAGE GALACTOSE N-ACETYL-GALACTOSAMINE SPECIFIC LECTIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC014811; AAH14811.1; -
KW Lectin.
SQ SEQUENCE 304 AA; 34622 MW; 2272ELADA2C0262A CRC64;

Query Match
Best Local Similarity 11.4%; Score 101.5; DB 11; Length 304;
Matches 36; Conservative 38; Mismatches 48; Indels 33; Gaps 7;

QY 24 LGIGLIVLLIIVLYVPLIITIKANSACRD-GIRAVMECRNVTHLTLOELTEAOKGF 81
DB 41 LGISLILVVSIVIG-----SONSOLRBDLGTIRALID--NTTSKIKAE-----F 83
QY 82 GQVEAQAATCNHTVMALMSLDAEKAQO-----KVEELEGET---TTINHKIQDAS 131
DB 84 QSLDSRAUDSEFGISLKYVDVHDHROELQAGRDLSOKYTSLESTLEKREQALKTLSLTL 143
QY 132 AEVELRRENOYLSRIADKKYPPSSODSSSAAP 166
DB 144 DHVQQLRKDLKALTOQLANLK-----NNGSEVACCP 174

RESULT 12

ID 09YHD5 PRELIMINARY; PRT; 826 AA.
AC 09YHD5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MYOSIN HEAVY CHAIN (FRAGMENT).
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana.

```

Query Match 11.3%; Score 100.5; DB 4; Length 506;
      Similarity 26.3%; Pred No. 1.9;
Match 41; Conservative 29; Mismatches 53; Indels 33; Gaps 7

QY 54 RDGLAAVME-----GRNVTHLL--QELTLEAKQGFQDVEAO--AATC-----NHTVMA 97
      |||:||||:||||:|:|:|
Db 8 RDSLHATNELLOVRRQSLTHILALOEELTRKVPSPDSLEPFTKCGSLNRWRKRYVA 67
      |||:||||:||||:|:|:|

QY 98 LMASIDAERKAGQCKVEELIGEITTLNKK-----LDASAE--VERLIRREN 141
      |||:||||:||||:|:|:|
Db 68 LMVQLKAELEHSDSVKOLKGVASHQEKVTSQSQEQALIORSLQDKAAEYGVERRMGANG 127
      |||:||||:||||:|:|:|

QY 142 QVLSVRIADKKYTPSSQSSSAAAPOLLIVLGLSA 177
      |||:||||:||||:|:|:|
Db 128 LQLELSRAQEARRRWQOQTAS--AEEDLRIRLVANVSS 162
      |||:||||:||||:|:|:|

```

RESULT	15	
	P79391	
ID	P79391	PRELIMINARY;
AC	P79391;	PRF; 270 AA.
DT	01-MAY-1997	(TREMBLrel. 03, Created)
DT	01-MAY-1997	(TREMBLrel. 03, Last sequence update)
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	LECTIN-LIKE OXIDIZED LDL RECEPTOR.	
OS	Bos taurus (bovine).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 18:42:42 ; Search time 11.94 Seconds
(without alignments)
583.712 Million cell updates/sec

Title: US-09-828-217-1
Perfect score: 889
Sequence: 1 MASTSYDYCRVPMEDGDKRC.....SSAAPQLLVLGLSALLQ 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	889	100.0	180	1	BST2_HUMAN	Q10589 homo sapien
2	102	11.5	550	1	KUCR_RAT	P10716 rattus norv
3	101	11.4	422	1	KICR_MOUSE	P05784 mus musculu
4	99.5	11.2	304	1	MUGL_MOUSE	P49300 mus musculu
5	99	11.1	962	1	VDP_HUMAN	O60763 homo sapien
6	98.5	11.1	1966	1	MTSB_CAEEL	P02566 caenorhabdi
7	98	11.0	1938	1	MYH4_RABIT	Q28641 oryctolagus
8	98	11.0	1938	1	MYSD_CAEEL	P02567 caenorhabdi
9	97	10.9	286	1	PV92_SCICO	P22312 sclara copr
10	97	10.9	959	1	VDP_RAT	P41542 rattus norv
11	96.5	10.9	893	1	VM92_CAEEL	P34531 caenorhabdi
12	96	10.8	286	1	P091_SCICO	P22311 sclara copr
13	95	10.7	1937	1	MYH8_HUMAN	P13535 homo sapien
14	95	10.7	1972	1	MYHB_RABIT	P35749 homo sapien
15	95	10.7	1972	1	MYHB_HUMAN	P35748 oryctolagus
16	94.5	10.6	1935	1	MYSS_CYPCA	O90339 cyprinus ca
17	94.5	10.6	4473	1	PLE1_CRIGR	O91155 cricetus
18	94	10.6	1084	1	MTSS_RABIT	P02563 oryctolagus
19	93.5	10.5	1938	1	MYH6_RAT	P05562 rattus norv
20	93.5	10.5	1939	1	MYH6_MESAU	P13539 mesocricetu
21	93	10.5	848	1	MYSP_DIRIM	P13592 dirofilaria
22	93	10.5	879	1	MYSP_ONCOV	O02171 onchocerca
23	93	10.5	880	1	MYSP_BRUMA	O01202 brugia mala
24	93	10.5	1972	1	MYHB_MOUSE	O08638 mus musculu
25	92.5	10.4	1509	1	MYTN_ACACA	P05659 acanthamoeb
26	92.5	10.4	1938	1	MYH6_MOUSE	O02566 mus musculu
27	92	10.3	882	1	MYSP_CAEEL	P10567 caenorhabdi
28	91.5	10.3	244	1	MYH7_PAPHA	P11778 papio hamad
29	91.5	10.3	428	1	PLT2_CARAU	O42305 carassius a
30	91.5	10.3	1102	1	MYSC_CHICK	P26616 gallus galli
31	91.5	10.3	1934	1	MYH7_MESAU	P13540 mesocricetu
32	91.5	10.3	1935	1	MYH7_HUMAN	P12883 homo sapien
33	91.5	10.3	1935	1	MYH7_RAT	P02564 rattus norv

34	91	10.2	407	1	M21_STRPY	P50468 streptococc
35	91	10.2	516	1	P54_ENTRC	P13692 enterococcu
36	90.5	10.2	692	1	MYS_PODCA	O05000 podocoryne
37	90.5	10.2	1846	1	MTSB_RAT	P70569 rattus norv
38	90.5	10.2	1938	1	MTSS_CAEEL	P13538 rattus norv
39	90.5	10.2	1939	1	MYH1_HUMAN	P12882 homo sapien
40	90.5	10.2	4687	1	PLE1_RAT	P30427 rattus norv
41	90	10.1	941	1	VDP_MOUSE	O92120 mus musculu
42	90	10.1	2116	1	MYSD_DICDI	P08799 dictyosteli
43	89.5	10.1	1938	1	MYHD_HUMAN	Q9UKX3 homo sapien
44	89.5	10.1	1939	1	MYHB_HUMAN	P13533 homo sapien
45	89	10.0	1978	1	MYHB_CHICK	P10587 gallus galli

ALIGNMENTS

RESULT	ID	STANDARD	PRT	180 AA.
BST2_HUMAN	Q10589			
AC	Q10589	STANDARD	PRT	180 AA.
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Bone marrow stromal antigen 2 (BST-2).			
GN	BST2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95331788; PubMed=7607676;			
RA	Ishtkawa J., Kaisho T., Tomizawa H., Lee B.O., Kobune Y.,			
RA	Inazawa J., Oritani K., Itoh M., Ochi T., Ishihara K., Hirano T.;			
RT	"Molecular cloning and chromosomal mapping of a bone marrow stromal			
RT	cell surface gene, BST2, that may be involved in pre-B-cell growth.";			
RL	Genomics 26:527-534(1995).			
CC	- FUNCTION: MAY BE INVOLVED IN PRE-B-CELL GROWTH.			
CC	- SUBCELLULAR LOCATION: Type II membrane protein.			
CC	- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER, LUNG, HEART			
CC	AND PLACENTA. LOWER LEVELS IN PANCREAS, KIDNEY, SKELETAL MUSCLE			
CC	AND BRAIN.			
CC	- DISEASE: MAY PLAY A ROLE IN B-CELL ACTIVATION IN RHEUMATOID			
CC	ARTHRITIS (RA).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: D28137; BAA05679.1; -			
DR	MIM: 600534; -			
KM	Transmembrane: Glycoprotein; Signal-anchor; Polymorphism.			
FT	DOMAIN 1 20			
FT	TRANSMEM			
FT	DOMAIN 21 48			
FT	DOMAIN 49 180			
FT	CARBOHYD 65 65			
FT	CARBOHYD 92 92			
FT	VARIANT 143 143			
FT	SEQUENCE 180 AA; 19769 MW; CAF52340D69061EE CRC64;			
SO	SEQUENCE			
QY	1 MASTSYDYCRVPMEDGDKRCLLIGILVLLIYLVPLITFTIKNSSEACRGRLRAY 60			
QY				

DB 1 MASTSDYCRVPMEDGDKRCKLLLGIGLIVLLIGVPLIIFTIKANSEACRDLIRAV 60
 QY 61 MECRNATVHLLOEELTEAOKGFQDVEAQAATCNHTVMAALMSILDAEKAOGKVEELEGEL 120
 DB 61 MECRNATVHLLOEELTEAOKGFQDVEAQAATCNHTVMAALMSILDAEKAOGKVEELEGEL 120
 QY 121 TTNHKLQDASAVEERLRRENOVLSVRIDAKKRYPPSSQDSSAAAPQLIVLIGLSALLQ 180
 DB 121 TTNHKLQDASAVEERLRRENOVLSVRIDAKKRYPPSSQDSSAAAPQLIVLIGLSALLQ 180
 RESULT 2
 KICR_RAT STANDARD: PRT: 550 AA.
 AC P10716;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 RT Kupfer cell receptor.
 KCLR.
 US Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 83-104.
 RX MEDLINE=88227939; PubMed=2836387;
 RA Hoyle G.W., Hill R.L.;
 RT "Molecular cloning and sequencing of a cDNA for a carbohydrate binding receptor unique to rat Kupfer cells.";
 RT J. Biol. Chem. 263:7487-7492(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91107689; PubMed=1846367;
 RA Hoyle G.W., Hill R.L.;
 RT "Structure of the gene for a carbohydrate-binding receptor unique to rat Kupfer cells.";
 RT J. Biol. Chem. 266:1850-1857(1991).
 CC -1- FUNCTION: RECEPTOR WITH AN AFFINITY FOR GALACTOSE AND FUCOSE.
 CC COULD BE INVOLVED IN ENDOCYTOSIS.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: KUPFER CELLS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL J03734; AAA41472.1; -
 DR EMBL M55532; AAA40892.1; -
 DR PIR: A28166; A28166.
 DR PIR: A38674; A38674.
 DR HSP: P20693; 1HLJ.
 DR InterPro: IPR000017; Syntaxin.
 DR InterPro: IPR001304; lectin_c.
 DR Pfam: PF00059; lectin_c; 1.
 DR SMART: SM00034; CLECT; 1.
 DR SMART: SM00503; SYN; 1.
 DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
 KM Receptor; Transmembrane; Glycoprotein; Lectin; Signal anchor;
 KM Endocytosis.
 FT DOMAIN 1 42
 FT TRANSMEM 43 69
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 70 550
 FT DOMAIN 438 538
 FT DISULFID 440 536
 FT DISULFID 516 528
 FT BY SIMILARITY.
 FT BY SIMILARITY.

FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61104 MW; 9358A6CF4C306270 CRC64;
 Query Match 11.5%; Score 102; DB 1; Length 550;
 Best local similarity 28.0%; Pred. No. 0.57;
 Matches 37; Conservative 20; Mismatches 61; Indels 14; Gaps 3;
 QY 46 IKANSEACRDLGRAVMECRNVTHLLOEELTEAOKGFQDVEAQAATCNHTVMAALMSILDAE 105
 DB 256 ISAEIOAMDQMDGMRAGE-----EMTSLKDELTFLTAQIONANGHLEQDTQIQGL 305
 QY 106 KAGCKKVEELEGELTTLNHKLQDASAVEERLRRENOVLSVRIDAKKRYPPSSQDSSAA 165
 DB 306 KAO-LKSTSLNSQILEVNGKLDSSRELOTLRD--LSDVSAKSNVQMLQSNLOKAK 361
 QY 166 POLIIVLIGLSA 177
 DB 362 AEVQSLKTGLEA 373
 RESULT 3
 KICR_MOUSE STANDARD: PRT: 422 AA.
 AC P05784; O61766;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Keratin, type I cytoskeletal 18 (Cytokeratin 18) (Cytokeratin endo B)
 DE (Keratin D).
 GN KRT18 OR KRT1-18 OR KERD.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89196920; PubMed=2467843;
 RA Ichinose Y., Morita T., Zhang F., Srimahosonggram S., Tondella M.L.C.,
 RA Matsumoto M., Nozaki M., Matsushiro A.;
 RT "Nucleotide sequence and structure of the mouse cytokeratin endo B gene.";
 RT Gene 70:85-95(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Teratocarcinoma;
 RA Alonso A., Weber T., Jorciano J.L.;
 RT "Cloning and characterization of keratin D, a murine endodermal cytoskeletal protein induced during in vitro differentiation of F9 teratocarcinoma cells.";
 RT Roux's Arch. Dev. Biol. 196:16-21(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86085876; PubMed=2416755;
 RA Singer P.A., Trevor K., Oshima R.G.;
 RT "Molecular cloning and characterization of the Endo B cytokeratin expressed in preimplantation mouse embryos.";
 RT J. Biol. Chem. 261:538-547(1986).
 RN [4]
 RP SEQUENCE OF 1-131 FROM N.A.
 RX MEDLINE=88255838; PubMed=2454868;
 RA Oshima R.G., Trevor K., Shevinsky L.H., Ryder O.A., Cecena G.;
 RT "Identification of the gene coding for the Endo B murine cytokeratin and its methylated, stable inactive state in mouse nonepithelial cells.";
 RT Genes Dev. 2:505-516(1988).
 CC -1- SUBUNIT: HETERODIMER OF TWO TYPE I AND TWO TYPE II KERATINS.
 CC KERATIN 18 ASSOCIATES WITH KERATIN 8.


```

CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFILAMENTAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M22832; AAA37552.1; -
DR EMBL: M36376; AAA39373.1; -
DR EMBL: M11686; AAA39390.1; -
DR EMBL: Y00217; CAA68365.1; -
DR PIR: A25621; A25621.
DR PIR: A28428; A28428.
DR PIR: J70406; J70406.
DR SWISS-2DPAGE: P05784; MOUSE.
DR MGI: 96692; Krt1-18.
DR InterPro: IPR001664; IF.
DR InterPro: IPR002957; Keratin_I.
DR Pfam: PF00038; filament; 1.
DR PRINTS: PR01248; TYPE1KERATIN.
DR PROSITE: PS00226; IF; 1.
DR Intermediate filament; Coiled coil; Keratin; Glycoprotein;
KW Acetylation.
FT INT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT DOMAIN 1 70 HEAD.
FT DOMAIN 1 71 ROD.
FT DOMAIN 380 422 TAIL.
FT DOMAIN 71 106 COIL 1A.
FT DOMAIN 107 124 LINKER 1.
FT DOMAIN 125 216 COIL 1B.
FT DOMAIN 217 240 LINKER 12.
FT DOMAIN 241 379 COIL 2.
FT SITE 263 263 STUTTER.
FT SITE 323 323 STUTTER.
FT CARBOHYD 30 30 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 31 31 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 49 49 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CONFLICT 133 133 L -> F (IN REF. 1).
FT CONFLICT 243 243 D -> N (IN REF. 2).
FT CONFLICT 252 252 A -> G (IN REF. 2).
SQ SEQUENCE 422 AA; 47373 MW; 4D5B0E9C7732FE2F3 CRC64;

```

```

ID MGL_MOUSE STANDARD; PRT; 304 AA.
AC P49300;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Macrophage asialoglycoprotein-binding protein (M-ASGP-BP) (Macrophage
DE galactose/N-acetylgalactosamine-specific lectin) (MGL).
GN MGL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HEN;
RX MEDLINE=92268032; PubMed=1587794;
RA Sato M., Kawakami K., Osawa T., Toyoshima S.;
RT "Molecular cloning and expression of cDNA encoding a galactose/N-
RT acetylgalactosamine-specific lectin on mouse tumoricidal
RT macrophages.";
RL J. Biochem. 111:331-336(1992).
RN [2]
RP SEQUENCE OF 102-120 AND 137-151.
RC STRAIN=C3H/HEN;
RX MEDLINE=89197865; PubMed=3241002;
RA Oda S., Sato M., Toyoshima S., Osawa T.;
RT "Purification and characterization of a lectin-like molecule specific
RT for galactose/N-acetyl-galactosamine from tumoricidal macrophages.";
RL J. Biochem. 104:600-605(1988).
CC -1- FUNCTION: RECOGNIZES TERMINAL GALACTOSE AND N-ACETYLGALACTOSAMINE
CC UNITS. MAY PARTICIPATE IN THE INTERACTION BETWEEN TUMORICIDAL
CC MACROPHAGES AND TUMOR CELLS.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- TISSUE SPECIFICITY: IS EXPRESSED ON THE SURFACE OF ACTIVATED
CC MACROPHAGES.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S36676; AAB22171.1; -
DR HSSP: P06734; IJKE.
DR MGI: 96975; MGL.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
KW Lectin; Glycoprotein; Transmembrane; Calcium; Signal-anchor.
FT DOMAIN 1 35 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 36 56 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 57 304 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 172 298 C-TYPE LECTIN (LONG FORM).
FT DISULFD 173 184 BY SIMILARITY.
FT DISULFD 201 296 BY SIMILARITY.
FT DISULFD 274 288 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 304 AA; 34596 MW; 3F79CD12C34F5BCC CRC64;

```

```

Query Match 11.2%; Score 99.5; DB 1; Length 304;
Best Local Similarity 23.2%; Pred. NO. 0.47;
Matches 36; Conservative 37; Mismatches 49; Indels 33; Gaps 7;

```

OY 24 LGIGLVLLIIVLIGVPLIIFIKANSEACRD-GLRAVMECRNVTYHLQQLTEAOKGF 81

Db 41 LGSLLLLVVSVIG-----SQNSQLRRDGLTTRATLD-NTTSKRIAE-----F 83
 QY 82 QDVEAQAATCNHTVMAIMASLDAEKAAGQ-----KKVELEGEI-----TTLNKKLQDAS 131
 Db 84 QSDSRSDSEKKGSSKXVEEDHROELQAGRDLSCKVTSLSESTVEKREQALKTDLSDLT 143
 QY 132 AEVERLRENOVLVRIADKKYPPSSODSSSAAP 166
 Db 144 DHVQLRKDKALCOLANLK-----NNGSEVACCP 174

RESULT 5

VDP_HUMAN STANDARD; PRT; 962 AA.

AC 060763;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE General vesicular transport factor p115 (transcytosis associated protein) (TAP) (Vesicle docking protein).
 GN VDP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PHOSPHORYLATION SITE SER-942.
 RX MEDLINE=98148093; PubMed=9478999;
 RA Sohda M., Misumi Y., Yano A., Takami N., Ikehara Y.;
 RT "Phosphorylation of the vesicle docking protein p115 regulates its association with the Golgi membrane."
 RL J. Biol. Chem. 273:5385-5388(1998).
 CC -1- FUNCTION: GENERAL VESICULAR TRANSPORT FACTOR REQUIRED FOR INTERCYSTERNAL TRANSPORT IN THE GOLGI STACK: IT IS REQUIRED FOR TRANSCYTOTIC FUSION AND/OR SUBSEQUENT BINDING OF THE VESICLES TO THE TARGET MEMBRANE. MAY WELL ACT AS A VESICULAR ANCHOR BY INTERACTING WITH THE TARGET MEMBRANE AND HOLDING THE VESICULAR AND TARGET MEMBRANES IN PROXIMITY (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN WHICH RECYCLES BETWEEN THE CYTOSOL AND THE GOLGI APPARATUS DURING INTERPHASE.
 CC -1- DOMAIN: COMPOSED OF A GLOBULAR HEAD, AN ELONGATED TAIL (COILED-COIL) AND A HIGHLY ACIDIC C-TERMINAL DOMAIN.
 CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-SPECIFIC MANNER; PHOSPHORYLATED IN INTERPHASE BUT NOT IN MITOTIC CELLS.
 CC DEPHOSPHORYLATED PROTEIN ASSOCIATES WITH THE GOLGI MEMBRANE; PHOSPHORYLATION PROMOTES DISSOCIATION.
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D86326; BAA25300.1; -
 DR MIM; 603344; -
 DR InterPro: IPR000225; Armadillo.
 DR PROSITE; PS50176; ARM_REPEAT; UNKNOWN_1.
 KW Transport; Protein transport; Golgi stack; Membrane; Coiled coil;
 KW Phosphorylation.
 FT DOMAIN 1 637 GLOBULAR HEAD.
 FT DOMAIN 638 930 COILED COIL (POTENTIAL).
 FT DOMAIN 935 962 ASP/GLU-RICH (ACIDIC).
 FT MOD_RES 942 942 ASP/PHOSPHORYLATION.
 FT MUTGEN 942 942 S->A: LOSS OF PHOSPHORYLATION.
 SQ SEQUENCE 962 AA; 107906 MW; 2E748F2C1BCB942 CRC64;

Query Match 11.1%; Score 99; DB 1; Length 962;
 Best Local Similarity 28.5%; Pred. No. 1.8;

Matches 35; Conservative 26; Mismatches 54; Indels 8; Gaps 3;
 QY 60 VMECRVYTHLLQOELTEAOKGFQDVE-AQAATCNHTVMAIMASLDAEKAAGQKKVELEEG 118
 Db 739 IEELKKNQELQSQLKEKSMIEMKSSQTSNGNESSAIVSARDE-----QVAEIKQ 792
 QY 119 EITTLNKKLQDASAEVERLRENO-VLSVRIADKKYPPSSODSSSAAPDQLIVLIGLSA 177
 Db 793 ELATLKSQLNSQSVETIKQTEKQELQTEAFPAKSVVEGETETIATRTDVEGRISA 852
 QY 178 LLQ 180
 Db 853 LLQ 855

RESULT 6

MYSB_CAEEL STANDARD; PRT; 1966 AA.

AC P02366;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain B (MHC B).
 GN UNC-54 OR MYO-4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83273600; PubMed=6576334;
 RA Karn J., Brenner S., Barnett L.;
 RT "Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy chain gene are not separated by introns."
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
 CC [2]
 CC SEQUENCE OF 850-1966 FROM N.A.
 RX MEDLINE=82272395; PubMed=7202124;
 RA McLachlan A.D., Karn J.;
 RT "Periodic charge distributions in the myosin rod amino acid sequence RT match cross-bridge spacings in muscle."
 RL Nature 299:226-231(1982).
 CC [3]
 CC SEQUENCE OF 1876-1966 FROM N.A.
 RX MEDLINE=83232892; PubMed=6571695;
 RA Willis N., Gesteland R.F., Karn J., Barnett L., Bolten S.,
 RA Waterston R.H.;
 RT "The genes sup-7 x and sup-5 III of C. elegans suppress amber nonsense mutations via altered transfer RNA."
 RL Cell 33:575-583(1983).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C.ELEGANS.
 CC -1- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: J01050; AAA28124.1; -;
DR EMBL: V01494; CAA24738.1; -;
DR PIR: A02992; MMKW.
DR HSSP: P08799; 1MND.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF02736; Myosin_N.1.
DR Pfam: PF01576; Myosin_tail.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head.1.
DR SMART: SM00242; MYSC.1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 850 COILED COIL (POTENTIAL).
FT DOMAIN 851 1966 ALPHA-HELICAL TAILPIECE (S2).
FT DOMAIN 851 1164 HINGE.
FT NP_BIND 1165 1966 LIGHT MEROMYOSIN (LMM).
FT NP_BIND 177 184 ATP (BY SIMILARITY).
FT DOMAIN 665 687 ACTIN-BINDING.
FT DOMAIN 769 783 ACTIN-BINDING.
FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 705 705 ALKYLATION (SH-1).
FT MOD_RES 715 715 ALKYLATION (SH-2).
FT CONFLICT 1337 1337 E -> R (IN REF. 2).
FT CONFLICT 1880 1880 I -> L (IN REF. 2).
SQ SEQUENCE 1966 AA; 225125 MW; B66F0B2FE27B67F CRC64;

Query Match 11.1%; Score 98.5; DB 1; Length 1966;
Best Local Similarity 24.7%; Pred. No. 4.2;
Matches 36; Conservative 21; Mismatches 58; Indels 31; Gaps 5;

OY 47 KANSEACDGLRAVMECRNVTLLDQELTEAGKGFQDYEAQAATCNHTVMALMASLDAEK 106
DB 1811 EEAFAALKGKVKVIAKLEQRYVELESLDEQRFQDANKNGRADRVRELQFQVDECK 1870
OY 107 AQ-----GCKVVELEGEITTLN-----HKLQDASEVERLRREN 141
DB 1871 KNEERLQDLIDKLOOKLKTKQKQVEEAE-ELANLMDQKYKQLTQLEDAEERAD--QAEN 1927
OY 142 QVLAVRI--ADKRYPPSSQDSSAA 164
DB 1928 SLSKMRSKRASASVAPGLQSSASAA 1953

RESULT 7
ID MYH4_RABIT STANDARD: PRT: 1938 AA.
AC Q28641.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, juvenile.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Skeletal muscle;
RA Maeda K., Hostinova E., Roesch-Kleinlauf A., Schuster H., Gasperik J.,
RA Wittinghofer A.;
RT "Isolation, sequencing of myosin heavy chain cDNA from rabbit
RT skeletal muscle and a novel cosynthesis of S-1 fragment with the
RT essential and regulatory light chains.";

RL submitted (Aug-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATTER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U32574; AAA74199.1; -;
DR HSSP: P08799; 1MND.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF02736; Myosin_N.1.
DR Pfam: PF01576; Myosin_tail.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head.1.
DR SMART: SM00015; IQ.1.
DR SMART: SM00242; MYSC.1.
DR PROSITE: PS50096; IQ.1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 783 MYOSIN HEAD-LIKE.
FT DOMAIN 784 813 IQ.
FT NP_BIND 842 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 658 680 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 35 35 METHYLATION (MONO-) (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 698 698 ALKYLATION (SH-1) (BY SIMILARITY).
FT MOD_RES 708 708 ALKYLATION (SH-2) (BY SIMILARITY).
SQ SEQUENCE 1938 AA; 223064 MW; D8A8A2E5B182626 CRC64;

Query Match 11.0%; Score 98; DB 1; Length 1938;
Best Local Similarity 26.1%; Pred. No. 4.6;
Matches 37; Conservative 20; Mismatches 45; Indels 40; Gaps 5;

OY 47 KANSEAC-----DGLRAVMECRNVTLLDQELTEAGKGFQDYEAQAATCNHTVMALM 99
DB 1366 KANSEVQMRTRYETDAIQRTELEBAKKRIARLQDAE---EHVEAVNAK----- 1414
OY 100 ASLDAEKAQGGKVELEGEITTLN-----HKLQDASEVERLR 138
DB 1415 ASLEKTKORLQNEVEDLMDIVERTNAACAALDKKQRPDKIIAEKKHKEETHAELASQ 1474
OY 139 RENQVLSRIAD-KRYPPSSQD 159
DB 1475 KRSRSISTEVFKVKNAYEESLD 1496

```

RESULT      8
MYSD_CAEEL  STANDARD; PRT; 1938 AA.
ID MYSD_CAEEL 019674;
AC P02567; Q19674;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain D (MHC D).
GN MYO-1 OR R06C7.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
ON NCBI_TaxID=6239;
RX MEDLINE=89178677; PubMed=2926820;
RC STRAIN-BRISTOL N2;
RN SEQUENCE FROM N.A.
RA MEDLINE=83273600; PubMed=6576334;
RX MEDLINE=83273600; PubMed=6576334;
RA Karn J., Brenner S., Barnett L.;
RT "Protein structural domains in the Caenorhabditis elegans unc-54
RT myosin heavy chain gene are not separated by introns.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
RN [3]
RP SEQUENCE OF 115-365 AND 1492-1763 FROM N.A.
RX MEDLINE=85201409; PubMed=3888374;
RA Karn J., Dobb N.J., Miller D.M.;
RT "Cloning nematode myosin genes.";
RL Cell Muscle Motil. 6:185-237(1985).
RN [4]
RP SEQUENCE FROM N.A.
RA Gardner A., McMurray A.;
RC STRAIN-BRISTOL N2;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (LMC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 26-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
CC C.ELEGANS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X08065; CAA30854.1; -
DR EMBL: M37232; AAA28119.1; -
DR EMBL: M37234; AAA28120.1; -
DR EMBL: 271266; CAA95848.1; -
DR EMBL: 271261; CAA95848.1; JOINED.
DR EMBL: 271261; CAA95806.1; -

```

```

DR EMBL: 271266; CAA95806.1; JOINED.
DR PIR: S02772; MKKW1.
DR HSSP: P08799; 1MDN.
DR WormRep: R06C7.10; CE06253.
DR InterPro: IPR004009; Myosin N.
DR InterPro: IPR002928; Myosin tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF02736; Myosin_N; 1.
DR Pfam: PF01576; Myosin_tail; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00242; Msc; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 845
FT DOMAIN 846 1938
FT DOMAIN 846 1170
FT DOMAIN 1171 1938
FT NE_BIND 177 184
FT DOMAIN 660 682
FT DOMAIN 764 778
FT MOD_RES 128 128
FT MOD_RES 700 700
FT MOD_RES 710 710
FT MOD_RES 94 94
FT CONFLICT 98 98
FT CONFLICT 377 377
FT CONFLICT 389 390
FT CONFLICT 391 391
FT CONFLICT 408 408
FT CONFLICT 474 474
FT CONFLICT 577 577
FT CONFLICT 681 681
FT CONFLICT 1373 1373
FT CONFLICT 1659 1659
SQ SEQUENCE 1938 AA; 223255 MW; 387399CB6344CP4 CRC64;
Query Match 11.0%; Score 98; DB 1; Length 1938;
Best Local Similarity 26.6%; Pred. No. 4.6;
Matches 34; Conservative 24; Mismatches 58; Indels 12; Gaps 4;
OY 51 EACRQDGLRAVME-----CRNYTHLQO--ELTEAO-KGFQDVEAQAATCNHTVALMASLD 103
DB 1324 KAMEDLHENGQERHACKNLEHEDQHELLEQINGKDIDQQLSRINSEISQWARYE 1383
OY 104 AEKAGCKAYEELEGETTINHLKLDASAVEYLRRENOYLSRIADKTYPSQSSSA 163
DB 1384 GEGVASEELEELEKRRQMRVMDLQELSA-----QNKVISLEKAKGLLAETEDARSD 1438
OY 164 AAPQLLIV 171
DB 1439 VDRHLTYI 1446
RESULT      9
P092_SCICO  STANDARD; PRT; 286 AA.
ID P092_SCICO 022312;
AC P22312;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PUFF II/9-2 protein precursor.
GN PUFF II/9-2.
OS Sciatara coprophila (Fungus gnat).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Sciarioidea;
OC Sciarioidea; Bradysia.
ON NCBI_TaxID=38358;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-6980;
 RX MEDLINE=90133907; PubMed=2614832;
 RA Dibratolomeis S.M., Gerbi S.A.;
 RT "Molecular characterization of DNA puff II/9A genes in Sclara
 RT copophila.";
 RL J. Mol. Biol. 210:531-540(1989).
 CC -I- MISCELLANEOUS: THE DNA PUFF II/9 PROTEINS HAVE A PROPOSED
 CC INTERMOLECULAR COILED COIL STRUCTURE WITH POSSIBLY INTERMOLECULAR
 CC DISULFIDE BRIDGES FORMED BY NUMEROUS CYSTEINE RESIDUES IN POSITION
 CC D OF THE HEPTAD REPEAT
 CC -I- SIMILARITY: 76% IDENTICAL TO THE PUFF II/9-1 PROTEIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC DR EMBL: X51679; CAA35982.1; -
 CC PIR: S07533; S07533.
 CC SIGNAL: coiled coil; Glycoprotein.
 CC FT CHAIN 1 19 OR 21 (POTENTIAL).
 CC FT DOMAIN 61 286 PUFF II/9-2 PROTEIN.
 CC FT CAROHD 156 156 HELICAL (POTENTIAL).
 CC FT SEQUENCE 286 AA; 32621 MW; 720AC8CCC22A869C CRC64;
 SQ

Query Match 10.9%; Score 97; DB 1; Length 286;
 Best Local Similarity 25.6%; Pred. No. 0.7;
 Matches 31; Conservative 24; Mismatches 36; Indels 30; Gaps 5;

OY 46 IKAASBACDGLRAVMEC-RNVTHL-----IQDELTEAKGFQDVEAQAATCNHTMAL 98
 DB 89 IKREKARQAKAEKAEKCKNTENLKEETQELKELAEKQAEKCKKELADCK----- 142
 OY 99 MASLAEKQAGKVEELEGTTLNHKQ-----DASAVERLRRE-----NOVLSV 146
 DB 143 -----KENAKLNTKEELNCTTLOELKRCGRGRERDLOCDLDECKKLNICNELLAC 197
 OY 147 R 147
 DB 198 R 198

RESULT 10

VDP_RAT
 VDP_RAT STANDARD: PRT; 959 AA.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE General vesicular transport factor p115 (Transcytosis associated
 DE protein) (TAP) (Vesicle docking protein).
 GN VDP
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95132632; PubMed=7831323;
 RA Saperstein S.K., Walter D.M., Grosvenor A.R., Heuser J.E.,
 RA Waters M.G.;
 RT "p115 is a general vesicular transport factor related to the yeast
 RT endoplasmic reticulum to Golgi transport factor Uosp.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:522-526(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver;

RX MEDLINE=95132633; PubMed=7831324;
 RA Barroso M., Nelson D.S., Szul E.;
 RT "Transcytosis-associated protein (TAP)/p115 is a general fusion
 RT factor required for binding of vesicles to acceptor membranes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:527-531(1995).
 CC -I- FUNCTION: GENERAL VESICULAR TRANSPORT FACTOR REQUIRED FOR
 CC INTERCISTERNAL TRANSPORT IN THE GOLGI STACK; IT IS REQUIRED FOR
 CC TRANSCYTIC FUSION AND/OR SUBSEQUENT BINDING OF THE VESICLES TO
 CC THE TARGET MEMBRANE. MAY WELL ACT AS A VESICULAR ANCHOR BY
 CC INTERACTING WITH THE TARGET MEMBRANE AND HOLDING THE VESICULAR
 CC AND TARGET MEMBRANES IN PROXIMITY.
 CC -I- SUBUNIT: HOMODIMER WITH TWO N-TERMINAL HEADS AND A C-TERMINAL
 CC COILED-COIL TAIL. DIMER FORMED BY PARALLEL ASSOCIATION OF THE
 CC TAILS.
 CC -I- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN WHICH RECYCLES
 CC BETWEEN THE CYTOSOL AND THE GOLGI APPARATUS DURING INTERPHASE.
 CC -I- DOMAIN: COMPOSED OF A GLOBULAR HEAD, AN ELONGATED TAIL (COILED-
 CC COIL) AND A HIGHLY ACIDIC C-TERMINAL DOMAIN.
 CC -I- PTM: PHOSPHORYLATED IN A CELL CYCLE-SPECIFIC MANNER;
 CC PHOSPHORYLATED IN INTERPHASE BUT NOT IN MITOTIC CELLS.
 CC DEPHOSPHORYLATED PROTEIN ASSOCIATES WITH THE GOLGI MEMBRANE;
 CC PHOSPHORYLATION PROMOTES DISSOCIATION (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE VDP/USO1/YB1047C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC DR EMBL: U14192; AAA62632.1; -
 CC DR EMBL: U15389; AAC52151.1; -
 CC DR InterPro: IPR000225; Armadillo.
 CC DR PROSITE: PS50176; ARM_REPEAT; UNKNOWN_1.
 CC KW Transport; Protein transport; Golgi stack; Membrane; Coiled coil;
 CC KW Phosphorylation.
 CC FT DOMAIN 1 637 GLOBULAR HEAD.
 CC FT DOMAIN 638 930 COILED COIL (POTENTIAL).
 CC FT DOMAIN 935 959 ASP/GLU-RICH (ACIDIC).
 CC FT MOD_RES 940 940 PHOSPHORYLATION (BY SIMILARITY).
 CC FT CONFLICT 591 591 S -> P (IN REF. 2).
 CC FT CONFLICT 658 658 M -> V (IN REF. 2).
 CC FT CONFLICT 816 816 S -> R (IN REF. 2).
 CC FT CONFLICT 873 873 A -> S (IN REF. 2).
 CC SQ SEQUENCE 959 AA; 107162 MW; 356394B48C7E003B CRC64;

Query Match 10.9%; Score 97; DB 1; Length 959;
 Best Local Similarity 30.1%; Pred. No. 2.6;
 Matches 37; Conservative 20; Mismatches 58; Indels 8; Gaps 3;

OY 60 VMECRNVTHLLOOELTEAKGFQDV-EAQAATCNHTMALMASLAEKQAGKVEELBG 118
 DB 739 IEIRSHOVLDSQLAEKTVTEENLRSSQVSGMSBDAATCSPRDE-----QVAELKQ 792
 OY 119 EITTLNHLKLDASAVERLRRENOVLSVRAD-KKYPPSSODSSSAADQLITVLGISA 177
 DB 793 ELSALKSQLCSQSLFTRQTEENSELQRAETLAKSPVEGSELYTAKTIDVEGRISA 852
 OY 178 LQ 180
 DB 853 LQ 855
 RESULT 11
 YM92_CABEL
 ID YM92_CABEL STANDARD: PRT; 893 AA.
 AC P34531; P34532; P34533;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)

```

DN      Hypothetical 100.0 kDa protein M01A8.2 in chromosome III.
GN      M01A8.2.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
OC      Rhabdilitidae; Pelodierinae; Caenorhabdilis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BRISTOL N2;
RX      MEDLINE=94150718; PubMed=7906398;
RA      Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA      Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA      Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA      Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA      Johnston L., Jones M., Kerhaw J., Kirsten J., Lalster N.,
RA      Latreille P., Lighting J., Lloyd C., Mortimore B., O'Callaghan M.,
RA      Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA      Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA      Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA      Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA      Woldman P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans."
RL      Nature 368:32-38(1994).
CC      -1 SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL information
CC      CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z27081; CAAB1607.1; -.
DR      PIR; S40998; S40998.
DR      PIR; S40999; S40999.
DR      WormPep; M01A8.2; CE03491.
DR      InterPro; IPR000938; CAP-Gly.
DR      Pfam; PF01302; CAP_GLY_1;
DR      PROSITE; PS00845; CAP_GLY_1;
DR      PROSITE; PSS0245; CAP_GLY_2;
KW      Hypothetical protein; Coiled coil.
FT      DOMAIN          39               81       CAP-GLY.
FT      DOMAIN          522            696       COILED COIL (POTENTIAL).
FT      DOMAIN          729            756       COILED COIL (POTENTIAL).
FT      SEQUENCE        893 AA; 99997 MW; 464F2962B36C28B1 CRC64;
Query Match           10.9%; Score 96.5; DB 1; Length 893;
Best Local Similarity 24.8%; Pred. No. 2.6;
Matches   32; Conservative 29; Mismatches 39; Indels 29; Gaps 4;
OY      48 ANSEACRGRLRVMECRNVTHLLQ--QELTEKQKFQDY----EAQAACNTVVAIMAS 101
DB      566 SNQGVYINHNANV-LSLOKTHTETIAEKNKEFFERNFEERRARRREAVCMNHRQKVAC 624
OY      102 IDAEKAGOKCKVEELE-----GEITTLNHKLQDASAEEVERLR 139
DB      625 IDEKISEAKGCEQLNDKKVLQALANDCHRRNQMLTKREISLOTALIEKSAEKELRKQ 684
OY      QY      140 ENQYLVSRI 148
DB      Db      685 KNQNLSTQV 693
RESULT  12
ID      PU91_SCICO STANDARD: PRT; 286 AA.
AC      P22311;
DT      01-AUG-1991 (Rel. 19, Created)
DT      01-AUG-1991 (Rel. 19, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE PUFF II/9-1 protein precursor.
GN II/9-1.
OS Sciaera coprophila (Fungus gnat).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Scleroidea;
OC Sclariidae; Bradyzia.
OX NCBI_TaxID=389358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6980;
RX MEDLINE=90133907; PubMed=2614832;
RA Diapiccolomeis S.M., Gerbi S.A.;
RT "Molecular characterization of DNA puff II/9A genes in Sciaera
  coprophila.";
RJ J. Mol. Biol. 210:531-540(1989).
CC -I- MISCELLANEOUS: THE DNA PUFF II/9 PROTEINS HAVE A PROPOSED
CC INTERMOLECULAR COILED COIL STRUCTURE WITH POSSIBLY INTERMOLECULAR
CC DISULFIDE BRIDGES FORMED BY NUMEROUS CYSTEINE RESIDUES IN POSITION
CC D OF THE HEPAID REPEAT.
CC -I- SIMILARITY: 76% IDENTICAL TO THE PUFF II/9-2 PROTEIN.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51680; CAA35983.1; -.
DR PIR: S07532; S07532.
KW Signal; Coiled coil; Glycoprotein.
FT SIGNAL 1 19 OR 21 (POTENTIAL).
FT CHAIN 20 286 PUFF II/9-1 PROTEIN.
FT DOMAIN 61 235 HELICAL (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GICNAC.....) (POTENTIAL).
SQ SEQUENCE 286 AA; 32034 MW; AA6A7B55F191BBD1 CRC64;

Query Match 10.8%; Score 96; DB 1; Length 286;
Best Local Similarity 30.0%; Pred. No. 0.84;
Matches 24; Conservative 16; Mismatches 36; Indels 4; Gaps 1;

QY 59 AVMECRNTVHLIQELTEKQKGFQDEVQATQNTHTVMAWMSLDAEKAQGGKVEELG 118
   111: : 111: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 102 ALCEQKNSSELTKEITQELKELAQTKQELANCKREA---LANCKAEVNAKLKITEELNC 157
   111: : 111: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 119 EITTLNHLKQDASAEVERLR 138
   111: : 111: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 158 TITQLEELQECRRAREDLQ 177

RESULT 13
MYH8_HUMAN
AC P13535; Q14910; STANDARD; PRT; 1937 AA.
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, perinatal (MYHC-perinatal).
GN MYH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90323631; PubMed=2373371;
RA Karsch-Mizrachi I., Feghali R., Shown T.B. Jr., Leinwand L.A.;
RT "Generation of a full-length human perinatal myosin heavy-chain-
  encoding cDNA.";
RL Gene 89:289-294(1990).

```

[2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=95324556; PubMed=7601129;
 RA Julian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,
 RA Stehman H.H., Rudenstein N.A.;
 RT "Characterization of a human perinatal myosin heavy-chain
 transcript.";
 RL Eur. J. Biochem. 230:1001-1006(1995).
 RN [3]
 RP SEQUENCE OF 502-1937 FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=90235862; PubMed=1691980;
 RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 myosin heavy chains.";
 RL Eur. J. Biochem. 189:55-65(1990).
 RN [4]
 RP SEQUENCE OF 860-1937 FROM N.A.
 RX MEDLINE=89234168; PubMed=2715179;
 RA Peghail R., Leinwand L.A.;
 RT "Molecular genetic characterization of a developmentally regulated
 human perinatal myosin heavy chain.";
 RL J. Cell Biol. 108:1791-1797(1989).
 RN [5]
 RP SEQUENCE OF 1-46 FROM N.A.
 RA Esser K., Tidhar A., Myszkowski M.;
 RT "Isolation and characterization of the human perinatal MHC promoter.";
 RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALINE LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, M36769; AAC17185.1; -;
 CC EMBL, 238133; CA86293.1; -;
 CC EMBL, X51593; AA35941.1; -;
 CC EMBL, M35250; AA36346.1; -;
 CC EMBL, AF067143; AAC21557.1; -;
 CC PIR, A30220; A30220.
 CC HSSP, P08799; 1LVK.
 CC MIM, 160741; -;
 CC InterPro, IPR000048; IQ.
 CC InterPro, IPR004009; Myosin_N.
 CC InterPro, IPR002928; Myosin_tail.
 CC InterPro, IPR001609; myosin_head.
 CC Pfam, PF00612; IQ; 1.
 CC Pfam, PF00063; myosin_head; 1.
 CC Pfam, PF02736; myosin_N; 1.
 CC Pfam, PF01576; myosin_tail; 1.
 CC PRINTS, PR00193; MYOSINHEAVY.
 CC PRODOM, PD000355; myosin_head; 1.

DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 KN Myosin: Muscle protein; coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding.
 FT DOMAIN 1 780
 FT MYOSIN HEAD-LIKE.
 FT IQ.
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 842 1937
 FT ATP.
 FT NP_BIND 181 188
 FT ACTIN-BINDING.
 FT DOMAIN 658 680
 FT ACTIN-BINDING.
 FT MOD_RES 760 774
 FT METHYLATION (TR1-) (POTENTIAL).
 FT MOD_RES 132 132
 FT ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 698 698
 FT ALKYLATION (SH-2) (POTENTIAL).
 FT MOD_RES 708 708
 FT A -> R (IN REF. 2).
 FT CONFLICT 15 15
 FT E -> Q (IN REF. 1 AND 4).
 FT CONFLICT 970 970
 FT M -> N (IN REF. 3).
 FT CONFLICT 1072 1072
 FT N -> H (IN REF. 1 AND 4).
 FT CONFLICT 1247 1247
 FT MC -> DG (IN REF. 3).
 FT CONFLICT 1251 1252
 FT E -> G (IN REF. 1 AND 4).
 FT CONFLICT 1261 1261
 FT K -> Q (IN REF. 1 AND 4).
 FT CONFLICT 1297 1297
 FT RY -> NT (IN REF. 3).
 FT CONFLICT 1378 1378
 FT EN -> AH (IN REF. 1 AND 4).
 FT CONFLICT 1504 1505
 FT E -> D (IN REF. 1 AND 4).
 FT CONFLICT 1847 1847
 FT D -> H (IN REF. 2).
 FT CONFLICT 1914 1914
 SO SEQUENCE 1937 AA; 222762 MW; A3EE2D151792E9E8 CRC64;
 Query Match 10.7%; Score 95; DB 1; Length 1937;
 Best Local Similarity 25.2%; Pred. No. 7.9;
 Matches 33; Conservative 21; Mismatches 59; Indels 18; Gaps 4;
 QY 47 KANSEACR-----DGLRAVMECRNVTHLDOELTEAQKGVDEAQAATCNHTYMAIM 99
 DB 1366 KANSEVAQWRKYEDPAIORTTELEAKKKRLQRLQEAHEHVAVNAKASLEKKORIQ 1425
 QY 100 -----ASLDAEKAQG-----OKKVELEGEITLNNKLDASAEVRLRENOVLSVRIA 149
 DB 1426 NEVEDIMLVERSNAAALADKKRNFDVLSBWKOKYETQAEHLASQKESRSISTELP 1485
 QY 150 D-KKYPSSOD 159
 DB 1486 KVKKNYEESLD 1496
 RESULT 14
 MYHB_HUMAN STANDARD; PRT; 1972 AA.
 AC P35749; O00396; P78422; O94944;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, smooth muscle isoform (SMMHC).
 GN MYH1 OR KIA0866.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94425270; PubMed=10493829;
 RA Loftus B.J., Kim U.-T., Sneddon V.P., Kalush F., Brandon R.,
 RA Fuhrman J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
 RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
 RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
 RT "Genome duplications and other features in 12 Mb of DNA sequence from
 human chromosome 16p and 16q.";
 RL Genomics 60:295-308(1999).
 RN [2]
 RP SEQUENCE OF 1-1266 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;

RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
RN [3]
RP SEQUENCE OF 885-1972 FROM N.A.
RX MEDLINE=93263189; PubMed=7684189;
RA Matsuo R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,
RA Yanagisawa M., Masaki T., Takao A.;
RT "human smooth muscle myosin heavy chain gene mapped to chromosomal
RT region 16q12.";
RL Am. J. Med. Genet. 46:61-67(1993).
RN [4]
RP SEQUENCE OF 1093-1972 FROM N.A.
RC TISSUE=Hippocampus;
RA Okajima K.;
CC Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE; EXPRESSED IN THE UMBILICAL
CC ARTERY, BLADDER, ESOPHAGUS AND TRACHEA.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- DISEASE: A CHROMOSOMAL REARRANGEMENT, KNOWN AS PERICENTRIC
CC INVERSION INV(16)(P13Q22), PRODUCES A FUSION PROTEIN THAT CONSISTS
CC OF THE 165 N-TERMINAL RESIDUES OF CBF-BETA (PEPB2) WITH THE TAIL
CC REGION OF MYH11. THIS REARRANGEMENT IS ASSOCIATED WITH ACUTE
CC MYELOID LEUKEMIA OF MAEO SUBTYPE.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF001548; AAC31665.1; -;
DR EMBL: U91323; AAC35212.1; -;
DR EMBL: AB020673; BAA74889.1; -;
DR EMBL: D10667; -. NOT_ANNOTATED_CDS.
DR EMBL: X69292; CAA49134.1; -;
DR HSP: P08799; IMMN.
DR MIM: 160745; -;
DR InterPro: IPR000048; IQ.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF01576; myosin_tail; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 2.
DR SMART: SM00242; MWC; 1.
DR PROSITE: PS50096; IQ; 1.
DR Myosin: Muscle protein: Coiled coil; Thick filament; Actin-binding;
KM Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KM Multigene family; Proto-oncogene; Chromosomal translocation.

FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
FT 786 IQ.
FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
FT NP_BIND 178 185 ATP (POTENTIAL).
FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 762 776 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 129 129 METHYLATION (TR1-) (POTENTIAL).
FT MOD_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 887 889 EER -> NSE (IN REF. 3).
FT CONFLICT 1263 1266 ELOS -> TSLF (IN REF. 2).
FT CONFLICT 1558 1558 T -> S (IN REF. 3).
FT CONFLICT 1610 1611 KQ -> NE (IN REF. 3).
FT CONFLICT 1786 1786 A -> S (IN REF. 4).
FT CONFLICT 1958 1958 T -> L (IN REF. 3).
SQ SEQUENCE 1972 AA; 227338 MW; 676658B2ACE1277 CRC64;

Query Match 10.7%; Score 95; DB 1; Length 1972;
Best local similarity 24.5%; Pred. No. 8;
Matches 27; Conservative 18; Mismatches 55; Indels 10; Gaps 1;

QY 53 CROSLAVMECRNVTHTLLODELFEAQKGFQDVEQAATCNHTYAAIASD----- 103
DB 1368 CSDEERARAEINDKVHKLQNEVESVTGMNEAEKAKIKAKDAVASSLODQOELLQEE 1327
QY 104 -AEKAGQKVEELEGETITLHNKLQDASAEVERRENOVLSYRIADK 152
DB 1328 TRQRLNVTSLRQLQEEFRNSLDQDLDEMEKAKONLEHISTINQLSDSK 1377

RESULT 15
MYHB_RABIT
ID MYHB_RABIT STANDARD; PRT; 1972 AA.
AC P35748;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMHMC).
GN MYH11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
ON NCBI_TaxID=9986;
RX MEDLINE=92073350; PubMed=1961735;
RA Babb J.P., Kelly C., Pertisamy M.;
RT "Characterization of a mammalian smooth muscle myosin heavy-chain
RT gene: complete nucleotide and protein coding sequence and analysis of
RT the 5' end of the gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; M77812; AAA31395.1; -.
DR PIR; A41604; A41604.
DR HSSP; P08799; IMMD.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PSS0096; IQ; 1.
Myosin; muscle protein; Coiled coil; thick filament; Actin-binding;
Calmodulin-binding; ATP-binding; Methylation; Alkylation;
Multi-gene family.
KW DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 1 807 IQ.
FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
FT NP_BIND 178 185 ATP (POTENTIAL).
FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 763 777 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 129 129 METHYLATION (SH-1) (POTENTIAL).
FT MOD_RES 701 701 ALKYLATION (SH-2) (POTENTIAL).
FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1972 AA; 227318 MW; 2061A224288D6A4C CRC64;

Query Match 10.7%; Score 95; DB 1; Length 1972;
Best Local Similarity 24.5%; Pred. No. 8;
Matches 27; Conservative 18; Mismatches 55; Indels 10; Gaps 1;
QY 53 CRDGLRAVVECRNVTHTLQOELTEAOKGFQDVEAOATCNHTVMALMASLD----- 103
DB 1268 CSDGERARAELENDKVKHKLQNEVESVTGMLSEAEKAIKIAKEVASTLSGSLDPTQELLQEE 1327
QY 104 -AEKAQCKKVEELEGETTTLNHLQDASAEEYERLRRENOVLSVRIADKK 152
DB 1328 TRKILNVSTKLROLEDERNSLQELDEMEKAKONLERHISTINIOLEDSK 1377

Search completed: July 18, 2002, 18:44:50
Job time: 128 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 18:40:57 ; Search time 16.22 seconds
(without alignments)

1066.343 Million cell updates/sec

Title: US-09-828-217-1

Perfect score: 889

Sequence: 1 MASTSYDYCRVPMEDGDKRC.....SSAAPQLITVLGLSALIQ 180

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	889	100.0	180	2	A56836 bone marrow stroma
2	111	12.5	778	2	T30430 hypothetical prote
3	108	12.1	423	2	keratin, type I, c
4	107.5	12.1	461	2	H84099 cell wall-binding
5	107	12.0	143	2	G83955 flagellar protein
6	102.5	11.5	1534	2	A56734 ribosome receptor
7	102	11.5	550	2	A28166 Kupfer cell recep
8	99.5	11.2	304	2	JX0209 lectin, galactose/
9	99	11.1	415	2	S35760 fctr protein precu
10	98.5	11.1	1963	1	MMKW transcytosis-assoc
11	98	11.0	959	2	A55913 myosin heavy chain
12	98	11.0	1938	1	MMKW1 myosin heavy chain
13	98	11.0	1938	2	A59293 myosin heavy chain
14	97	10.9	286	2	S07533 hypothetical prote
15	96.5	10.9	597	2	S40998 hypothetical prote
16	96.5	10.9	893	2	G88551 protein M01A8.2 (l
17	96	10.8	286	2	S07532 puff II/9-1 protei
18	95.5	10.7	173	2	S76705 hypothetical prote
19	95	10.7	1937	2	I38055 myosin heavy chain
20	95	10.7	1972	1	A41604 myosin heavy chain
21	94.5	10.6	1388	2	S74245 serine/threonine-s
22	94	10.6	676	2	S00084 myosin heavy chain
23	93.5	10.5	140	2	H64629 hypothetical prote
24	93.5	10.5	1938	1	S06005 myosin alpha heavy
25	93.5	10.5	1939	2	I48175 myosin heavy chain
26	93	10.5	359	2	I50712 TOP AP - chicken
27	93	10.5	848	2	A44972 paramyosin - nemat
28	93	10.5	879	2	A46575 paramyosin - nemat
29	93	10.5	1938	2	JC5421 smooth muscle. myos

30	93	10.5	1972	2	JC5420 smooth muscle myos
31	92.5	10.4	392	2	G95258 secreted 45 kd pro
32	92.5	10.4	392	2	B98124 general stress pro
33	92.5	10.4	1509	1	A27224 myosin heavy chain
34	92.5	10.4	1938	2	I49464 alpha cardiac myos
35	92	10.3	746	2	T47237 myosin II heavy ch
36	92	10.3	866	2	S04027 paramyosin - Caeno
37	92	10.3	872	2	T19296 hypothetical prote
38	91.5	10.3	244	2	I36913 beta-myosin heavy
39	91.5	10.3	764	2	I51302 myosin heavy chain
40	91.5	10.3	1039	2	S18199 myosin heavy chain
41	91.5	10.3	1934	2	I48153 myosin heavy chain
42	91.5	10.3	1935	1	A37102 myosin beta heavy
43	91.5	10.3	1935	1	S06006 myosin beta heavy
44	91.5	10.3	1939	1	A46762 myosin alpha heavy
45	91.5	10.3	2442	2	T08621 centrosome associa

ALIGNMENTS

```

RESULT 1
A56836
bone marrow stromal cell surface protein BST-2 - human
C:Species: Homo sapiens (man)
C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 21-Jul-2000
C:Accession: A56836
R:Rishikawa, J.; Kaisho, T.; Tomizawa, H.; Lee, B.O.; Kobune, Y.; Inazawa, J.; Oriant
Genomics 26, 527-534, 1995
A:title: Molecular cloning and chromosomal mapping of a bone marrow stromal cell surf
A:Reference number: A56836; MUID:95331788
A:Accession: A56836
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-180 <180>
A:Cross-references: GB:D28137; MID:9457563; PIDN:BAA05679.1; PID:9506861
A:Gene: GDB:BST2
A:Cross-references: GDB:409946; OMTM:600534
A:Map position: 19p13.2-19p13.2
C:Keywords: transmembrane protein

Query Match
Best Local Similarity 100.0%; Score 889; DB 2; Length 180;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGIIVLITVILGVPLIIFTIKANSEACRDLRAV 60
|||||
Db 1 MASTSYDYCRVPMEDGDKRCKLLGIGIIVLITVILGVPLIIFTIKANSEACRDLRAV 60
|||||

QY 61 MECRNVTHTLLOEFLTEAGKGFQDVEAQAATCNHTVMAIASLDAEKAQCKKVEELEG 120
|||||
Db 61 MECRNVTHTLLOEFLTEAGKGFQDVEAQAATCNHTVMAIASLDAEKAQCKKVEELEG 120
|||||

QY 121 TITNHTLQDASAEVRLRENOVLSVRIDKRYPSQSSSSAAAPQLITVLGLSALIQ 180
|||||
Db 121 TITNHTLQDASAEVRLRENOVLSVRIDKRYPSQSSSSAAAPQLITVLGLSALIQ 180
|||||

RESULT 2
hypothetical protein ORF82 - Lymantria dispar nuclear polyhedrosis virus
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T30430
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; R
Virology 253, 17-34, 1999
A:title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria
A:Reference number: Z20836; MUID:99124785
A:Accession: T30430
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

```

A:Residues: 1-778 <KUZ>
A:Cross-references: EMBL:AF081810; PIDN:AACT0268.1

Query Match

12.5%: Score 111; DB 2; Length 778;
Best Local Similarity 32.5%: Pred. No. 0.48;

Matches 38; Conservative 22; Mismatches 47; Indels 10; Gaps 4;

OY 46 IKANSEACRDGLRAVMECR-NVTHLLOEELTEAOKGFQDVEAQAATCNHVMALMASIDA 104
DB 520 LKAQSEINRD-LQAKAEQADANARLQAEIDSLKR-----AESDAADLRNRAQJLEAEAS 574
OY 105 EKAQOKKVELEGEITTLNHNKLODASAEYERLRRENOVLSVRIADKKYPPSSODSS 161
DB 575 GAADLQNRIRARLEAESSGLTRRIQESAAEVYALRRKEDLERFPAAS-----AAQDVS 627

RESULT 3

9463

Atlatin, type I, cytoskeletal - mouse

N:Alternate names: endo B cytokeratin; keratin D

C:Species: Mus musculus (house mouse)

C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 10-Dec-1999

C:Accession: I59463; A25621; A28428; J00406

R:Alonso, A.; Weber, T.; Jorcano, J.L.

Roux's Arch. Dev. Biol. 196, 16-21, 1987

A:Title: Cloning and characterization of keratin D, a murine endodermal cytoskeletal p

A:Reference number: I59463

A:Accession: I59463

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-423 <RES>

A:Cross-references: GB:M36376; NID:g198587; PIDN:AAA9373.1; PID:g293682

R:Singler, P.A.; Trevor, K.; Oshima, R.G.

J. Biol. Chem. 261, 538-547, 1986

A:Title: Molecular cloning and characterization of the endo B cytokeratin expressed in F

A:Reference number: A25621; M0ID:86085876

A:Accession: A25621

A:Molecule type: mRNA

A:Residues: 1-243, 'D', 245-252, 'A', 254-423 <SIN>

A:Cross-references: GB:M1686; NID:g198620; PIDN:AAA9390.1; PID:g293685

R:oshima, R.G.; Trevor, K.; Shevinsky, L.H.; Ryder, O.A.; Cecena, G.

A:Title: Identification of the gene coding for the endo B murine cytokeratin and its me

A:Reference number: A28428; M0ID:88255838

A:Accession: A28428

A:Molecule type: DNA

A:Residues: 1-133 <OSH>

A:Cross-references: GB:Y00217; NID:g50842; PIDN:CAA68365.1; PID:g50843

R:Ichinose, Y.; Morita, T.; Zhang, F.; Srimalhasongcram, S.; Tondella, M.L.C.; Matsumoto, Gene 70, 85-95, 1988

A:Title: Nucleotide sequence and structure of the mouse cytokeratin endo gene.

A:Reference number: J00406; M0ID:89196920

A:Accession: J00406

A:Molecule type: DNA

A:Residues: 1-133, 'F', 135-243, 'D', 245-252, 'A', 254-423 <ICH>

A:Cross-references: GB:M28832; NID:g340757; PIDN:AAA37552.1; PID:g532610

C:Genetics:

A:Gene: endoB; KERD

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; intermediate filament

Query Match 12.1%: Score 108; DB 2; Length 423;
Best Local Similarity 28.0%: Pred. No. 0.44;

Matches 52; Conservative 28; Mismatches 56; Indels 50; Gaps 8;

OY 44 FTIKANSE-ACRD-----GLRAVMECRNVTHL-----LQOELTEAOKG-----FOD 83
DB 157 FAVKYETELAMRQSVSDHIGLRKRVYDNTITRLQLETEALKEKELLFKKHHHEEVYGS 216
OY 84 VEAQAATCNHTV-----MALMASIDAE-KAOGKKVEELE-----GEIT 121

DB 217 LEAOIASSGLTVEVDAPKSDLSKIMANIRAOYEALQOKRREEDKXWSQOIESTTVYF 276

OY 122 FLNHNKLODASAEYERLRRENOVLSVRIADKKYPPSSODSS-----SAAPOLLIVL 173

DB 277 TKSAEIRDAETTLTELRRLQLEIDIDSMKNONINLESLGVEARYKAKMEQLNGVL 336

OY 174 GLSALL 179
DB 337 HLESEL 342

RESULT 4

884099

cell wall-binding protein BH3600 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: H84099

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; M0ID:20512582; PMID:11058132

A:Accession: H84099

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-461 <STO>

A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA807319.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3600

Query Match 12.1%: Score 107.5; DB 2; Length 461;
Best Local Similarity 26.1%: Pred. No. 0.52;

Matches 31; Conservative 28; Mismatches 55; Indels 5; Gaps 2;

OY 33 IIVILGVPLIIFTIKANSEACRDGLRAVMECRNVTHLLOEELTEAOKGFQDVEAQAATCN 92
DB 5 ISLVAAAGLTFSLFSQSSIEDA-KANSLSIQDISDVOKERQEKQKTEAEEL----- 59
OY 93 HTVMALMASIDAEKAGOKKVELEGEITTLNHNKLODASAEYERLRRENOVLSVRIADK 151
DB 60 KEVEKELGDTIATERIDEKVEETSGKIOEKREIEVQAEIELEKQIELEIRIAER 118

RESULT 5

683955

flagellar protein required for flagellar formation flil [imported] - Bacillus halodur

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: G83955

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; M0ID:20512582; PMID:11058132

A:Accession: G83955

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-143 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA806166.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: flil

Query Match 12.0%: Score 107; DB 2; Length 143;
Best Local Similarity 22.1%: Pred. No. 0.17;

Matches 31; Conservative 34; Mismatches 49; Indels 26; Gaps 4;

OY 23 LIGIGLVLLIIVLGVPLIIFTIKANSEACRDGLRAVME-----RNVTHLLOQEL 74
DB 6 LVNIMLIIIVLTLVGVAALIFVNFNNEDEQREPTIDELIAQSYETEIEITTLISNDF 65

QY 75 TEA-----OKGFQDVEAQAATCNHTVAMLSIDAEGAOKKVELEGTEITTLNKK 126
 Db 66 VRARLHVDRNALQEVQKRFQYNNIIIRSLAGMDASQSLSGAIGIKLEAQ----- 118
 QY 127 LODASAEVERLRRENOVL 146
 Db 119 LOD--DINALMORGSVYKI 135

RESULT 6
 A56734
 ribosome receptor, 180k - dog
 C:Species: Canis lupus familiaris (dog)
 C:Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 08-Oct-1999
 C:Accession: A56734
 R:Maner, E.E., Sun, Y., Savitz, A.J., Meyer, D.I.,
 J. Cell Biol. 130, 29-39, 1995
 A:Title: Functional characterization of the 180-kD ribosome receptor in vivo.
 A:Reference number: A56734; MUID:95310363
 A:Accession: A56734
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1534 <MAN>
 A:Cross-references: GB:X87224; NID:g984113; PIDN:CAA60676.1; PID:g984114
 C:Keywords: endoplasmic reticulum; membrane protein; protein biosynthesis
 F:198-743/Region: 10-residue repeats (N-Q-G-K-K-A-E-G-A-P)

Query Match 11.5%; Score 102.5; DB 2; Length 1534;
 Best Local Similarity 28.6%; Pred. No. 4.4;
 Matches 26; Conservative 22; Mismatches 42; Indels 1; Gaps 1;

QY 55 DGLRVMGCRVNTHTLLOEELTEAOKGFQDVEAQAATCNHTVAMLSIDAEGAOKKVE 114
 Db 847 DAAVAKSKLRVNEKELAEKAAAGKAVKQIVAREOETTVAVARIEASYREHVKEVO 906
 QY 115 ELEGEITTLNKKLQDA-SAEVERLRRENOVL 144
 Db 907 QLOGKIRTLQEDLENGPNTQLARLQGENSIL 937

RESULT 7
 A28166
 Kupffer cell receptor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 05-Nov-1999
 C:Accession: A28166; A28166
 R:Hoyle, G.W., Hill, R.L.
 J. Biol. Chem. 266, 1850-1857, 1991
 A:Title: Structure of the gene for a carbohydrate-binding receptor unique to rat Kupffer
 A:Reference number: A38674; MUID:91107689
 A:Accession: A38674
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-550 <HO2>
 A:Cross-references: GB:M5532; NID:g203362; PIDN:AAA40892.1; PTD:g203363
 R:Hoyle, G.W., Hill, R.L.
 J. Biol. Chem. 263, 7487-7492, 1988
 A:Title: Molecular cloning and sequencing of a cDNA for a carbohydrate binding receptor
 A:Reference number: A28166; MUID:88227939
 A:Accession: A28166
 A:Molecule type: mRNA
 A:Residues: 1-550 <HOV>
 A:Cross-references: GB:J03734; NID:g205050; PIDN:AAA41472.1; PTD:g205051
 C:Superfamily: C-type lectin homology
 C:Keywords: transmembrane protein
 F:412-536/Domain: C-type lectin homology <LCH>

Query Match 11.5%; Score 102; DB 2; Length 550;
 Best Local Similarity 28.0%; Pred. No. 1.7;
 Matches 37; Conservative 20; Mismatches 61; Indels 14; Gaps 3;

QY 46 IKANSEACRDIIRAVMECRVNTHTLLOEELTEAOKGFQDVEAQAATCNHTVAMLSIDA 105
 Db 256 ISAEIQAMRDMQRRAGE-----EMTSIKNDLELTQIQNANHLEQDTQIOGL 305
 QY 106 KAOGKVEELEGETTLNKKLQDASAEVERLRRENOVL 146
 Db 306 KAO-LKSTSSINSQIEVNVGKIKDSRELQTLRRD---LSDVSAKSNVQMQLQAK 361
 QY 166 POLLIVILGLSA 177
 Db 362 AEVOSLKTGLEA 373

RESULT 8
 JX0209
 lectin, galactose/N-acetylglactosamine-specific - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Aug-1999
 C:Accession: JX0209; PX0009
 R:Sato, M.; Kawakami, K.; Osawa, T.; Toyoshima, S.
 J. Biochem. 111, 331-336, 1992
 A:Title: Molecular cloning and expression of cDNA encoding a galactose/N-acetylgalact
 A:Reference number: JX0209; MUID:92268032
 A:Accession: JX0209
 A:Molecule type: mRNA
 A:Residues: 1-304 <SAT>
 A:Cross-references: GB:S36676; NID:g249360; PIDN:AAB22171.1; PID:g249361
 R:Oda, S.; Sato, M.; Toyoshima, S.; Osawa, T.
 J. Biochem. 104, 600-605, 1988
 A:Title: Purification and characterization of a lectin-like molecule specific for gal
 A:Reference number: PX0009; MUID:89197865
 A:Accession: PX0009
 A:Molecule type: protein
 A:Residues: 102-120;137,X',139-151 <ODA>
 C:Superfamily: hepatic lectin; C-type lectin homology
 C:Keywords: glycoprotein; lectin; macrophage; transmembrane protein
 F:36-61/Domain: transmembrane status predicted <TRA>
 F:173-296/Domain: C-type lectin homology <LCH>
 F:74,166/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 11.2%; Score 99.5; DB 2; Length 304;
 Best Local Similarity 23.2%; Pred. No. 1.4;
 Matches 36; Conservative 37; Mismatches 49; Indels 33; Gaps 7;

QY 24 LGIGLIVLITVIGVPLIIFTIKANSEACRD--GLRVMGCRVNTHTLLOEELTEAOKGF 81
 Db 41 LGLSLILLVVSIVIG-----SONQLRDLGTLRATLD--NTSRKIAE-----F 83
 QY 82 ODVEAQAATCNHTVAMLSIDAEGAOKG-----KVEELEGETL---TTLNKKLQDAS 131
 Db 84 OSLDSRADSPFEKGISLKYVDHROELQAGDLSQKVTSLSTVEKRBQALKTDLSDLT 143
 QY 132 AEVERLRRENOVL 146
 Db 144 DHVQQLRDLKALTLQGLANLK-----NMGSEVACP 174

RESULT 9
 S35760
 forA protein precursor - Streptococcus pyogenes
 C:Species: Streptococcus pyogenes
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-Aug-1999
 C:Accession: S35760; A42711
 R:Podbielski, A.
 submitted to the EMBL Data Library, November 1992
 A:Reference number: S35760
 A:Accession: S35760
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-415 <POD>
 A:Cross-references: EMBL:X69324; NID:g311759; PIDN:CAA49165.1; PTD:g311760
 R:Haanes, E.J.; Heath, D.G.; Cleary, P.P.

DB 1439 VDRHLTVI 1446

RESULT 13

A59293

skeletal myosin heavy chain - domestic rabbit

C:Species: *Oryctolagus cuniculus*

C:Date: 09-Jun-2000 #sequenceRevision 09-Jun-2000 #text-change 08-Sep-2000

C:Accession: A59293

R:Macada, K.; Hostilnova, E.; Roesc Kleinkauf, A.; Schuster, H.; Gasperik, J.; Witting

submitted to GenBank, July 1995

A:Description: Isolation, sequencing of myosin heavy chain cDNA from rabbit skeletal

A:Reference number: A59293

A:Accession: A59293

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1938 <MAE>

A:Cross-references: GB:U52574; NID:g940232; PIDN:AA74199.1; PID:g940233

A:Experimental source: strain New Zealand White; cell type skeletal muscle fiber type

C:Genetics:

A:Gene: MHC

C:Superfamily: myosin heavy chain; myosin motor domain homology

F:89-769/Domain: myosin motor domain homology <MMO>

RESULT 14
S07533
puff II/9A-2 protein precursor - fungus gnat (Sciara coprophila)
C:Species: Sciara coprophila
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Jun-2000
C:Accession: S07533
R:DiBartolomeis, S.M.; Gerbi, S.A.
J. Mol. Biol. 210, 531-540, 1989
A:Title: Molecular characterisation of DNA puff II/9A genes in Sciara coprophila
A:Reference number: S07532; M01D:90133907
A:Accession: S07533
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-286 <DIB>
A:Cross-references: GB:X51679; NID:g10113; PID:g1405812
C:Genetics:
A:Map position: II/9A
C:Keywords: coiled coil; glycoprotein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-286/Product: puff II/9A protein #status predicted <MAT>
F:156/Binding site: carboxylate (Asn) (covalent) #status predicted

	Query Match	Score	97:	DB	Length	286:
Best Local Similarity	25.6%	Pred	No. 2.1,			
Matches	31;	Conservative	24;	Mismatches	36;	Indels 30; Gaps 5.
QY	46	IANSFACDGLRAYMEC-RNVTHL-----LQQELTEAKGFQDYEAQAATCNHTVMAL	98			
Db	89	LKRREKRAAEALKCECKNTENLKETEIOKLLEIAEOKRQLERKKKTLACK-----	142			

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 18:38:22 ; Search time 12.96 Seconds
(Without alignments)
339.244 Million cell updates/sec

Title: US-09-828-217-1

Perfect score: 889

Sequence: 1 NASTSYDYCRPMEDGDKRC.....SSAAPOLLIVLGLSALLQ 180

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters:

231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/p10data/2/1aa/5A_COMB.pep:*
2: /cgn2_6/p10data/2/1aa/5B_COMB.pep:*
3: /cgn2_6/p10data/2/1aa/6A_COMB.pep:*
4: /cgn2_6/p10data/2/1aa/6B_COMB.pep:*
5: /cgn2_6/p10data/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/p10data/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889	100.0	180	US-08-624-650-1	Sequence 1, Appl1
2	99.5	11.2	270	US-09-035-095-4	Sequence 4, Appl1
3	99.5	11.2	270	US-08-809-494A-2	Sequence 2, Appl1
4	99.5	11.2	270	US-09-352-302-2	Sequence 4, Appl1
5	99.5	11.2	273	US-08-809-494A-4	Sequence 4, Appl1
6	99.5	11.2	273	US-09-352-302-4	Sequence 4, Appl1
7	95	10.7	885	US-08-533-306A-4	Sequence 4, Appl1
8	95	10.7	885	US-08-742-923A-4	Sequence 4, Appl1
9	93.5	10.5	1886	US-08-938-105-3	Sequence 4, Appl1
10	92.5	10.4	292	US-08-688-342-4	Sequence 4, Appl1
11	92.5	10.4	292	US-09-113-788-4	Sequence 4, Appl1
12	89.5	10.1	316	US-09-111-470-4	Sequence 4, Appl1
13	89.5	10.1	1939	US-09-310-187A-1	Sequence 1, Appl1
14	88	9.9	288	US-08-312-949-4	Sequence 4, Appl1
15	88	9.9	288	US-08-446-201-4	Sequence 4, Appl1
16	88	9.9	619	US-08-465-746-2	Sequence 2, Appl1
17	88	9.9	619	US-08-214-164-2	Sequence 2, Appl1
18	88	9.9	619	US-08-467-852A-3	Sequence 3, Appl1
19	88	9.9	619	US-08-246-636-3	Sequence 2, Appl1
20	88	9.9	619	US-08-247-491A-3	Sequence 3, Appl1
21	88	9.9	619	US-08-319-795-2	Sequence 2, Appl1
22	88	9.9	619	US-08-468-985-2	Sequence 2, Appl1
23	88	9.9	619	US-08-312-949-2	Sequence 2, Appl1
24	88	9.9	648	US-08-072-070-2	Sequence 2, Appl1
25	88	9.9	648	US-08-469-434-2	Sequence 2, Appl1
26	88	9.9	648	US-08-214-222-2	Sequence 2, Appl1
27	88	9.9	648	US-08-467-852A-2	Sequence 2, Appl1

28	88	9.9	648	2	US-08-468-718-2	Sequence 2, Appl1
29	88	9.9	648	2	US-08-247-491A-2	Sequence 2, Appl1
30	88	9.9	648	1	US-08-446-201-3	Sequence 2, Appl1
31	88	9.9	605	1	US-08-127-499A-23	Sequence 23, Appl1
32	88	9.9	655	1	US-08-482-847-23	Sequence 23, Appl1
33	87.5	9.8	289	1	US-08-072-070-4	Sequence 4, Appl1
34	87.5	9.8	289	1	US-08-469-434-4	Sequence 4, Appl1
35	87.5	9.8	289	1	US-08-214-222-4	Sequence 4, Appl1
36	87.5	9.8	289	2	US-08-467-852A-5	Sequence 5, Appl1
37	87.5	9.8	289	2	US-08-468-718-4	Sequence 4, Appl1
38	87.5	9.8	289	2	US-08-247-491A-5	Sequence 5, Appl1
39	87.5	9.8	477	1	US-08-402-217A-3	Sequence 3, Appl1
40	87.5	9.8	477	1	US-08-700-178-3	Sequence 3, Appl1
41	87.5	9.8	477	3	US-08-995-654-3	Sequence 3, Appl1
42	87.5	9.8	667	4	US-09-071-709-9	Sequence 9, Appl1
43	85.5	9.6	459	4	US-09-071-709-1	Sequence 1, Appl1
44	85	9.6	344	6	5210183-2	Patent No. 5210183
45	85	9.6	683	6	5210183-3	Patent No. 5210183

ALIGNMENTS

RESULT 1
US-08-624-650-1
Sequence 1, Application US/08624650
Patent No. 5914252
GENERAL INFORMATION:
APPLICANT: HIRANO, TOSHIO
APPLICANT: KAISHO, TSUNEYASU
TITLE OF INVENTION: MEMBRANE PROTEIN POLYPEPTIDE HAVING
TITLE OF INVENTION: PRE-B CELL GROWTH-SUPPORTING ABILITY AND A GENE THEREOF
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,650
FILING DATE: 22-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP94/01732
FILING DATE: 14-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-281622
FILING DATE: 15-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 7625-001-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-624-650-1

Query Match 100.0%; Score 889; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 4.3e-88;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSDYCRVPEMDGDKCKLLGIGIIVLITVLITVGLPIFTIRKANEACRDLRAV 60
DB 1 MASTSDYCRVPEMDGDKCKLLGIGIIVLITVLITVGLPIFTIRKANEACRDLRAV 60
QY 61 MECRNVTLLQOELTEAOKGFQDVEAQAATCNHTVMAIASLDAERAKGCKKVEELGET 120
DB 61 MECRNVTLLQOELTEAOKGFQDVEAQAATCNHTVMAIASLDAERAKGCKKVEELGET 120
QY 121 TTLNHLKODASAEVRLRENOVLSVRIADKKYPPSSQDSSAAPOLLIVLIGISALIQ 180
DB 121 TTLNHLKODASAEVRLRENOVLSVRIADKKYPPSSQDSSAAPOLLIVLIGISALIQ 180

RESULT 2

US-09-055-095-4
Sequence 4, Application US/09055095

GENERAL INFORMATION:
PATENT NO. 5945308
APPLICANT: Tan9, Y. Tom
APPLICANT: Cortley, Chandra
APPLICANT: Cortley, Neil C.
APPLICANT: Sather, Susan
TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,095
FILING DATE: Filed Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0500 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1902982
US-09-055-095-4

Query Match 11.2%; Score 99.5; DB 2; Length 270;
Best Local Similarity 25.9%; Pred. No. 0.0068;
Matches 36; Conservative 25; Mismatches 41; Indels 37; Gaps 5;

QY 23 LIGIGILVLLIIVLIGVPLIIFIRKANEACRDLRAVMECRNVTHLLOELTEAOKGFQ 82
DB 40 VLCLGLVTVLIIILQLSOVSDLIKQ-----ANTH--QEDILEGO---- 80

QY 83 DVEAQAATCNHTVMAIASLDAERAKGCKKVEELGETITTLNHLKODASAEVRLRENO 142
DB 81 -----ILAQRSEKS-AQESQKELKEMITLHLAKHDKESKILMELHROVL 124

QY 143 VLSVRIADKKY--PSSOD 159
DB 125 NLQEVLEKANYSGPCPD 143

RESULT 3

US-08-809-494A-2
Sequence 2, Application US/08809494A

GENERAL INFORMATION:
PATENT NO. 5962260
APPLICANT: Masaki, Tatsuya
APPLICANT: Sawamura, Tomoo
TITLE OF INVENTION: Modified Low-Density Lipoprotein
TITLE OF INVENTION: Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAlay Fisher Nissen Goldberg & Kiel
STREET: 261 Madison Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10016-2391

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/809,494A
FILING DATE: 24-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-321705
FILING DATE: 30-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-214206
FILING DATE: 31-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Goldberg, Jules E.
REGISTRATION NUMBER: 24408
REFERENCE/DOCKET NUMBER: JG-YY-4363PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 986-4090
TELEFAX: 212 818-9479
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-809-494A-2

Query Match 11.2%; Score 99.5; DB 2; Length 270;
Best Local Similarity 25.9%; Pred. No. 0.0068;
Matches 36; Conservative 25; Mismatches 41; Indels 37; Gaps 5;

QY 23 LIGIGILVLLIIVLIGVPLIIFIRKANEACRDLRAVMECRNVTHLLOELTEAOKGFQ 82
DB 40 VLCLGLVTVLIIILQLSOVSDLIKQ-----ANTH--QEDILEGO---- 80

QY 83 DVEAQAATCNHTVMAIASLDAERAKGCKKVEELGETITTLNHLKODASAEVRLRENO 142
DB 81 -----ILAQRSEKS-AQESQKELKEMITLHLAKHDKESKILMELHROVL 124

QY 143 VLSVRIADKKY--PSSOD 159
DB 125 NLQEVLEKANYSGPCPD 143

```

      APPLICANT: Masaki, Tomoo  

      TITLE OF INVENTION: Modified Low-Density Lipoprotein Receptor  

      NUMBER OF SEQUENCES: 8  

      CORRESPONDENCE ADDRESS:  

        ADDRESSEE: Mcaulay Fisher Nissen Goldberg & Kiel  

          STREET: 261 Madison Avenue  

            CITY : New York  

              STATE: NY  

                COUNTRY: USA  

                  ZIP: 10016-2391  

                    COMPUTER READABLE FORM:  

                      MEDIUM TYPE: Floppy disk  

                        OPERATING SYSTEM: IBM PC compatible  

                          SOFTWARE: Patentln Release #1.0, Version #1.30  

                            CURRENT APPLICATION DATA:  

                              APPLICATION NUMBER: US/08/809,494A  

                                FILING DATE: 24-MAR-1997  

                                  CLASSIFICATION: 435  

                                    PRIOR APPLICATION DATA:  

                                      APPLICATION NUMBER: JP 6-321705  

                                        FILING DATE: 30-NOV-1994  

                                          PRIOR APPLICATION DATA:  

                                            APPLICATION NUMBER: JP 7-214206  

                                              FILING DATE: 31-JUL-1995  

                                                ATTORNEY/AGENT INFORMATION:  

                                                  NAME: Goldberg, Jules E.  

                                                    REGISTRATION NUMBER: 24408  

                                                      REFERENCE/DOCKET NUMBER: JS-YJ-4363PCT  

                                                        TELECOMMUNICATION INFORMATION:  

                                                          TELEPHONE: 212 986-4090  

                                                            TELEFAX: 212 818-9479  

                                                              INFORMATION FOR SEQ ID NO: 4:  

                                                                SEQUENCE CHARACTERISTICS:  

                                                                  LENGTH: 273 amino acids  

                                                                    TYPE: amino acid  

                                                                      TOPOLOGY: linear  

                                                                        MOLECULE TYPE: protein  

                                                                          US-08-809-494A-4
```

```

Query Match               11.2%; Score 99.5; DB 2; Length 273;  

Best Local Similarity     25.9%; Pred. No. 0.0069;  

Matches   36; Conservative    25; Mismatches    41; Indels    37; Gaps    5;
```

```
QY       23 ILIGIIILIVLITGVPLTIETIKANSEACDGLRAVMESCNVTNHLIQOELTEAKGFG 82  
         ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB       43 VLCGLIETVLTLTLLQLSOVSDLTKQQ-----ANTH--QEEDLEGG---- 83  
QY       83 DVEAQATCNTHTVALMSLDPAEKAGOKRKEELEGETTTLNHKIODASAEVERLRRENO 142  
Db       84 -----ILLQRSEKS-AQSQKELKMETLTAAHKDESKKIEMELHRNTL 127  
QY       143 VLSTRIADKKY--PPSOD 159  
         |:::|::|::|::|::|::|  
Db       128 NLQEVLIKANYSGPCPD 146
```

```
RESULT           6  
US-09-352-302-4  
Sequence 4, Application US/09352302  
Patent No. 6197937  
GENERAL INFORMATION:  
APPLICANT: Sawamura, Tatsuya  
APPLICANT: Masaki, Tomoo  
TITLE OF INVENTION: Modified Low-Density Lipoprotein Receptor  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mcaulay Fisher Nissen Goldberg & Kiel  
STREET: 261 Madison Avenue  
CITY: New York
```

QY	53	CRDGIATMECRNTHLLQOELTEKGFQDYEAQAATCNTHVMA	LM	SLD	-----	103
Db	215	CSDGRRARALNDKYKHLQNEEVSATGALNNEGGAIKIAKYVAS	SS	LOD	QTO	ELDDE 274
QY	104	-AEKAGQKVEELEGETITLNLHKLODASAERYLRRENOVL	SV	RIAD	KK 152	
Db	275	TROKLANSTKRLQREERNSSLODQDDEEMKAKONERHIST	LN	IDS	OSK 324	

RESULT 8
US-08-742-923A-4
; Sequence 4, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.

```

1 STATE: MI
2 COUNTRY: USA
3 ZIP: 48303
4
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM PC compatible
8 OPERATING SYSTEM: PC-DOS/MS-DOS
9 SOFTWARE: Patent In Release #1.0, Version #1.25
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/08/742,923A
12 FILING DATE: No. 5869611ember 1, 1996
13 CLASSIFICATION: 435
14 ATTORNEY/AGENT INFORMATION:
15 NAME: Smith, Dean F.
16 REGISTRATION NUMBER: 36683
17 REFERENCE/DOCKET NUMBER: 2115-00869DVC
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (810) 641-1600
20 TELEFAX: (810) 641-0270
21 INFORMATION FOR SEQ ID NO: 4:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 885 amino acids

```

```

; GENERAL INFORMATION:
;
; APPLICANT: Au-Young, Janice

```


RESULT 15
US-08-446-201-4
; Sequence 4, Application US/08446201B

Search completed: July 18, 2002, 18:42:58
Job time: 276 sec

THIS PAGE BLANK (USPTO)